Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/GB04/005462

International filing date: 23 December 2004 (23.12.2004)

Document type: Certified copy of priority document

Document details: Country/Office: GB
Number: 0329681.1
Filing date: 23 December 2003 (23.12.2003)

Date of receipt at the International Bureau: 21 February 2005 (21.02.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

PCT/GB2004/005462

The Patent Office

PATENTS • DESIGNS • COPYRIGHT • TRADE MARKS

INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed R. Mahoney

Dated 9 February 2005

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

THE PATENT OFFICE
SN
23 DEC 2003
RECEIVED BY FAX

23DEC03 E861442 D02846
P01/7700 0.00-0329681.1 ACCOUNT CHA

1/77

The Patent Office

Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

1.	Your reference	DELE / P28655GB
2.	Patent application number *(The Patent Office will fill in this part)*	23 DEC 2003 0329681.1
3.	Full name, address and postcode of the or of each applicant *(underline all surnames)*	Delta Biotechnology ~~Ltd~~ LIMITED Castle Court 59 Castle Boulevard Nottingham NG7 1FD United Kingdom
	05677729001 Patents ADP number *(if you know it)*	
	If the applicant is a corporate body, give the country/state of its incorporation	United Kingdom
4.	Title of the invention	GENE EXPRESSION TECHNIQUE
5.	Name of your agent *(if you have one)* "Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*	ERIC POTTER CLARKSON PARK VIEW HOUSE 58 THE ROPEWALK NOTTINGHAM NG1 5DD
	Patents ADP number *(if you know it)*	1305010
6.	If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number	Country / Priority application number *(if you know it)* / Date of filing *(day / month / year)*
7.	If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application	Number of earlier application / Date of filing *(day / month / year)*
8.	Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:* a) *any applicant named in part 3 is not an inventor; or* b) *there is an inventor who is not named as an applicant, or* c) *any named applicant is a corporate body.* *See note (d))*	YES

PL 1/77 24.04

0089811 23-Dec-03 12:30

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	71
Claims(s)	5
Abstract	1
Drawing(s)	25

SN

10. If you are also filing in any of the following, state how many against each item.

Priority Documents	0
Translations of priority documents	0
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	NO
Request for preliminary examination and search *(Patents Form 9/77)*	YES
Request for substantive examination *(Patents Form 10/77)*	NO
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature Eric Potter Clarkson
ERIC POTTER CLARKSON

Date
23 December 2003

12. Name and daytime telephone number of person to contact in the United Kingdom

0115 9552211

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 01645 500505.*
b) *Write your answers in capital letters using black ink or you may type them.*
c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*
d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*
e) *Once you have filled in the form you must remember to sign and date it.*
f) *For details of the fee and ways to pay please contact the Patent Office.*

GENE EXPRESSION TECHNIQUE

FIELD OF THE INVENTION

The present application relates to gene expression techniques.

BACKGROUND OF THE INVENTION

The class of proteins known as chaperones have been defined by Hartl (1996, *Nature*, **381**, 571-580) as a protein that binds to and stabilises an otherwise unstable conformer of another protein and, by controlled binding and release, facilitates its correct fate *in vivo*, be it folding, oligomeric assembly, transport to a particular subcellular compartment, or disposal by degradation.

BiP (also known as GRP78, Ig heavy chain binding protein and Kar2p in yeast) is an abundant ~70kDa chaperone of the hsp 70 family, resident in the endoplasmic reticulum (ER), which amongst other functions, serves to assist in transport in the secretory system and fold proteins.

Protein disulphide isomerase (PDI) is a chaperone protein, resident in the ER that is involved in the catalysis of disulphide bond formation during the post-translational processing of proteins.

Studies of the secretion of both native and foreign proteins have been shown that transit from the ER to the Golgi is the rate-limiting step. Evidence points to a transient association of the BiP with normal proteins and a more stable interaction with mutant or misfolded forms of a protein. As a result, BiP may play a dual role in solubilising folding precursors and

preventing the transport of unfolded and unassembled proteins. Robinson and Wittrup, 1995, *Biotechnol. Prog.* **11**, 171-177, have examined the effect of foreign protein secretion on BiP (Kar2p) and PDI protein levels in *Saccharomyces cerevisiae* and found that prolonged constitutive expression of foreign secreted proteins reduces soluble BiP and PDI to levels undetectable by Western analysis. The lowering of ER chaperone and foldase levels as a consequence of heterologous protein secretion has important implications for attempts to improve yeast expression/secretion systems.

Expression of chaperones is regulated by a number of mechanisms, including the unfolded protein response (UPR).

Using recombinant techniques, multiple PDI gene copies has been shown to increase PDI protein levels in a host cell (Farquhar *et al*, 1991, *Gene*, **108**, 81-89).

Co-expression of the gene encoding PDI and a gene encoding a heterologous disulphide-bonded protein was first suggested in WO 93/25676, published on 23 December 1993, as a means of increasing the production of the heterologous protein. WO 93/25676 reports that the recombinant expression of antistasin and tick anticoagulant protein can be increased by co-expression with PDI.

This strategy has been exploited to increase the recombinant expression of other types of protein.

Robinson *et al*, 1994, *Bio/Technology*, 12, 381-384 reported that a recombinant additional PDI gene copy in *Saccharomyces cerevisiae* could

be used to increase the recombinant expression of human platelet derived growth factor (PDGF) B homodimer by ten-fold and *Schizosacharomyces pombe* acid phosphatase by four-fold.

Hayano *et al*, 1995, *FEBS Letters*, 377, 505-511 described the co-expression of human lysozyme and PDI in yeast. Increases of around 30-60% in functional lysozyme production and secretion were observed.

Shusta *et al*, 1998, *Nature Biotechnology*, 16, 773-777 reported that the recombinant expression of single-chain antibody fragments (scFv) in *Saccharomyces cerevisiae* could be increased by between 2-8 fold by over-expressing PDI in the host cell.

Bao & Fukuhara, 2001, *Gene*, 272, 103-110 reported that the expression and secretion of recombinant human serum albumin (rHSA) in the yeast *Kluyveromyces lactis* could be increased by 15-fold or more by co-expression with an additional recombinant copy of the yeast PDI gene (*KlPDI1*).

In order to produce co-transformed yeast comprising both a PDI gene and a gene for a heterologous protein, WO 93/25676 taught that the two genes could be chromosomally integrated; one could be chromosomally integrated and one present on a plasmid; each gene could be introduced on a different plasmid; or both genes could be introduced on the same plasmid. WO 93/25676 exemplified expression of antistasin from the plasmid pKH4α2 in yeast strains having a chromosomally integrated additional copy of a PDI gene (Examples 16 and 17); expression of antistasin from the vector K991 with an additional PDI gene copy being present on a multicopy yeast shuttle vector named YEp24 (Botstein *et al*, 1979, *Gene*, 8, 17-24) (Example 20);

and expression of both the antistasin and the PDI genes from the yeast shuttle vector pC1/1 (Rosenberg *et al*, 1984, *Nature*, **312**, 77-80) under control of the GAL10 and GAL1 promoters, respectively. Indeed, Robinson and Wittrup, 1995, *op. cit.*, also used the GAL1 – GAL10 intergenic region to express erythropoietin and concluded that production yeast strains for the secretion of heterologous proteins should be constructed using tightly repressible, inducible promoters, otherwise the negative effects of sustained secretion (i.e. lowered detectable BiP and PDI) would be dominant after the many generations of cell growth required to fill a large-scale fermenter.

Subsequent work in the field has identified chromosomal integration of transgenes as the key to maximising recombinant protein production.

Robinson *et al*, 1994, *op. cit.*, obtained the observed increases in expression of PDGF and *S. pombe* acid phosphatase using an additional chromosomally integrated PDI gene copy. Robinson *et al* reported that attempts to use the multi-copy 2µm expression vector to increase PDI protein levels had had a detrimental effect on heterologous protein secretion.

Hayano *et al*, 1995, *op. cit.* described the introduction of genes for human lysozyme and PDI into a yeast host each on a separate linearised integration vector, thereby to bring about chromosomal integration.

Shusta *et al*, 1998, *op. cit.*, reported that in yeast systems, the choice between integration of a transgene into the host chromosome versus the use of episomal expression vectors can greatly affect secretion and, with reference to Parekh & Wittrup, 1997, *Biotechnol. Prog.*, **13**, 117-122, that

stable integration of the scFv gene into the host chromosome using a δ integration vector was superior to the use of a 2μm-based expression plasmid. Parekh & Wittrup, *op. cit.*, had previously taught that the expression of bovine pancreatic trypsin inhibitor (BPTI) was increased by an order of magnitude using a δ integration vector rather than a 2μm-based expression plasmid. The 2μm-based expression plasmid was said to be counter-productive for the production of heterologous secreted protein.

Bao & Fukuhara, 2001, *op. cit.*, reported that "It was first thought that the *KlPDI1* gene might be directly introduced into the multi-copy vector that carried the rHSA expression cassette. However, such constructs were found to severely affect yeast growth and plasmid stability. This confirmed our previous finding that the *KlPDI1* gene on a multi-copy vector was detrimental to growth of *K. lactis* cells (Bao *et al*, 2000)". Bao *et al*, 2000, *Yeast*, 16, 329-341, as referred to in the above-quoted passage of Bao & Fukuhara, reported that the *KlPDI1* gene had been introduced into *K. lactis* an a multi-copy plasmid, pKan707, and that the presence of the plasmid caused the strain to grow poorly. Bao *et al* concluded that over-expression of the *KlPDI1* gene was toxic to *K. lactis* cells. In the light of the earlier findings in Bao *et al*, Bao & Fukuhara chose to introduce a single duplication of *KlPDI1* on the host chromosome.

Against this background, we have surprisingly demonstrated that, contrary to the suggestions in the prior art, when the genes for a chaperone protein and a heterologous protein are co-expressed on a 2μm-family multi-copy plasmid in yeast, the production of the heterologous protein is substantially increased.

DESCRIPTION OF THE INVENTION

A first aspect of the present invention provides a method for producing heterologous protein comprising:

(a) providing a host cell comprising a 2µm-family plasmid, the plasmid comprising a gene encoding a protein comprising the sequence of a chaperone protein and a gene encoding a heterologous protein;

(b) culturing the host cell in a culture medium under conditions that allow the expression of the gene encoding the chaperone protein and the gene encoding a heterologous protein;

(c) purifying the thus expressed heterologous protein from the culture medium; and

(d) optionally, lyophilising the thus purified protein.

In one embodiment, step (c) purifies the thus expressed heterologous protein to a pharmaceutically acceptable level of purity.

Preferably, the method further comprises the step of formulating the purified heterologous protein with a carrier or diluent, such as a pharmaceutically acceptable carrier or diluent and optionally presenting the thus formulated protein in a unit dosage form.

A second aspect of the present invention provides for the use of a 2µm-family plasmid as an expression vector to increase the production of a

fungal (preferably yeast) or vertebrate heterologous protein by providing a gene encoding the heterologous protein and a gene encoding a protein comprising the sequence of a chaperone protein on the same 2µm-family plasmid.

A third aspect of the present invention provides a 2µm-family plasmid comprising a gene encoding a protein comprising the sequence of a chaperone protein and a gene encoding a heterologous protein, wherein if the plasmid is based on the 2µm plasmid then it is a disintegration vector.

A fourth aspect of the invention provides a host cell comprising a plasmid as defined above.

The present invention relates to recombinantly modified versions of 2µm-family plasmids.

Certain closely related species of budding yeast have been shown to contain naturally occurring circular double stranded DNA plasmids. These plasmids, collectively termed 2µm-family plasmids, include pSR1, pSB3 and pSB4 from *Zygosaccharomyces rouxii* (formerly classified as *Zygosaccharomyces bisporus*), plasmids pSB1 and pSB2 from *Zygosaccharomyces bailii*, plasmid pSM1 from *Zygosaccharomyces fermentati*, plasmid pKD1 from *Kluyveromyces drosphilarum*, an un-named plasmid from *Pichia membranaefaciens* (hereinafter "pPM1") and the 2µm plasmid and variants (such as Scp1, Scp2 and Scp3) from *Saccharomyces cerevisiae* (Volkert, *et al.*, 1989, *Microbiological Reviews*, 53, 299; Murray *et al.*, 1988, *J. Mol. Biol.* 200, 601; Painting, *et al.*, 1984, *J. Applied Bacteriology*, 56, 331).

As a family of plasmids these molecules share a series of common features in that they typically possess two inverted repeats on opposite sides of the plasmid, have a similar size around 6-kbp (range 4757 to 6615-bp), three open reading frames, one of which encodes for a site specific recombinase (*FLP*) and an autonomously replicating sequence (*ARS*), also known as an origin of replication (*ori*), located close to the end of one of the inverted repeats. (Futcher, 1988, *Yeast*, 4, 27; Murray *et al.*, *op. cit.*, and Toh-e *et al.*, 1986, *Basic Life Sci.* **40**, 425). Despite their lack of discernible DNA sequence homology, their shared molecular architecture and the conservation of function of the three open reading frames have demonstrated a common ancestral link between the family members.

Whilst any of the above naturally occurring 2μm-family plasmids can be used in the present invention, this invention is not limited to the use of naturally occurring 2μm-family plasmids. For the purposes of this invention, a 2μm-family plasmid is as described below.

A 2μm-family plasmid is a circular, double stranded, DNA plasmid. It is typically small, such as between 3,000 to 10,000 bp, preferably between 4,500 to 7000 bp, excluding recombinantly inserted sequences.

A 2μm-family plasmid typically comprises at least three open reading frames ("ORFs") that each encodes a protein that functions in the stable maintenance of the 2μm-family plasmid as a multicopy plasmid. The proteins encoded by the three ORFs can be designated FLP, REP1 and REP2. Where a 2μm-family plasmid comprises not all three of the ORFs encoding FLP, REP1 and REP2 then ORFs encoding the missing protein(s) should be supplied in *trans*, either on another plasmid or by chromosomal integration.

A "FLP" protein is a protein capable of catalysing the site-specific recombination between inverted repeat sequences recognised by FLP. The inverted repeat sequences are termed FLP recombination target (FRT) sites and each is typically present as part of a larger inverted repeat (see below). Preferred FLP proteins comprise the sequence of the FLP proteins encoded by one of plasmids pSR1, pSB1, pSB2, pSB3, pSB4, pSM1, pKD1, pPM1 and the 2µm plasmid, for example as described in Volkert *et al*, *op. cit.*, Murray *et al*, *op. cit.*, and Painting *et al.*, *op. cit.* Variants and fragments of these FLP proteins are also included in the present invention. "Fragments" and "variants" are those which retain the ability of the native protein to catalyse the site-specific recombination between the same FRT sequences. Such variants and fragments will usually have at least 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, or more, homology with an FLP protein encoded by one of plasmids pSR1, pSB1, pSB2, pSB3, pSB4, pSM1, pKD1, pPM1 and the 2µm plasmid. Different FLP proteins can have different FRT sequence specificities. A typical FRT site may comprise a core nucleotide sequence flanked by inverted repeat sequences. In the 2µm plasmid, the FRT core sequence is 8 nucleotides in length and the flanking inverted repeat sequences are 13 nucleotides in length (Volkert *et al*, *op. cit.*). However the FRT site recognised by any given FLP protein may be different to the 2µm plasmid FRT site.

REP1 and REP2 are proteins involved in the partitioning of plasmid copies during cell division, and may also have a role in the regulation of FLP expression. Considerable sequence divergence has been observed between REP1 proteins from different 2µm-family plasmids, whereas no sequence alignment is possible between REP2 proteins derived from different 2µm-family plasmids. Preferred REP1 and REP2 proteins comprise the sequence

of the REP1 and REP2 proteins encoded by one of plasmids pSR1, pSB1, pSB2, pSB3, pSB4, pSM1, pKD1, pPM1 and the 2μm plasmid, for example as described in Volkert *et al, op. cit.*, Murray *et al, op. cit.*, and Painting *et al, op. cit.* Variants and fragments of these REP1 and REP2 proteins are also included in the present invention. "Fragments" and "variants" of REP1 and REP2 are those which, when encoded by the plasmid in place of the native ORF, do not substantially disrupt the stable multicopy maintenance of the plasmid within a suitable yeast population. Such variants and fragments of REP1 and REP2 will usually have at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, or more, homology with a REP1 and REP2 protein, respectively, as encoded by one of plasmids pSR1, pSB1, pSB2, pSB3, pSB4, pSM1, pKD1, pPM1 and the 2μm plasmid.

The REP1 and REP2 proteins encoded by the ORFs on the plasmid must be compatible. It is preferred that the REP1 and REP2 proteins have the sequences of REP1 and REP2 proteins encoded by the same naturally occurring 2μm-family plasmid, such as pSR1, pSB1, pSB2, pSB3, pSB4, pSM1, pKD1, pPM1 and the 2μm plasmid, or variant or fragments thereof.

A 2μm-family plasmid typically comprises two inverted repeat sequences. The inverted repeats may be any size, so long as they each contain an FRT site (see above). The inverted repeats are typically highly homologous. They may share greater than 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% or more sequence identity. In a preferred embodiment they are identical. Typically the inverted repeats are each between 200 to 1000 bp in length. Preferred inverted repeat sequences may each have a length of from 200 to 300 bp, 300 to 400 bp, 400 to 500 bp, 500 to 600 bp, 600 to 700 bp, 700 to 800 bp, 800 to 900 bp, or 900 to 1000 bp. Particularly preferred inverted repeats are those of the plasmids pSR1 (959

bp), pSB1 (675 bp), pSB2 (477 bp), pSB3 (391 bp), pSM1 (352 bp), pKD1 (346 bp), the 2μm plasmid (599 bp), pSB4 or pPM1.

The sequences of the inverted repeats may be varied. However, the sequences of the FRT site in each inverted repeat should be compatible with the specificity of the FLP protein encoded by the plasmid, thereby to enable the encoded FLP protein to act to catalyse the site-specific recombination between the inverted repeat sequences of the plasmid. Recombination between inverted repeat sequences (and thus the ability of the FLP protein to recognise the FRT sites with the plasmid) can be determined by methods known in the art. For example, a plasmid in a yeast cell under conditions that favour FLP expression can be assayed for changes in the restriction profile of the plasmid which would result from a change in the orientation of a region of the plasmid relative to another region of the plasmid. The detection of changes in restriction profile indicate that the FLP protein is able to recognise the FRT sites in the plasmid and therefore that the FRT site in each inverted repeat are compatible with the specificity of the FLP protein encoded by the plasmid.

In a particularly preferred embodiment, the sequences of inverted repeats, including the FRT sites, are derived from the same 2μm-family plasmid as the ORF encoding the FLP protein, such as pSR1, pSB1, pSB2, pSB3, pSB4, pSM1, pKD1, pPM1 or the 2μm plasmid.

The inverted repeats are typically positioned with the 2μm-family plasmid such that the two regions defined between the inverted repeats (e.g. such as defined as UL and US in the 2μm plasmid) are of approximately similar size, excluding exogenously introduced sequences such as transgenes. For example, one of the two regions may have a length equivalent to at least

40%, 50%, 60%, 70%, 80%, 90%, 95% or more, up to 100%, of the length of the other region.

A 2μm-family plasmid typically comprises the ORF that encodes FLP and one inverted repeat (arbitrarily termed "IR1" to distinguish it from the other inverted repeat mentioned in the next paragraph) juxtaposed in such a manner that IR1 occurs at the distal end of the FLP ORF, without any intervening coding sequence, for example as seen in the 2μm plasmid. By "distal end" in this context we mean the end of the FLP ORF opposite to the end from which the promoter initiates its transcription. In a preferred embodiment, the distal end of the FLP ORF overlaps with IR1.

A 2μm-family plasmid typically comprises the ORF that encodes REP2 and the other inverted repeat (arbitrarily termed "IR2" to distinguish it from IR1 mentioned in the previous paragraph) juxtaposed in such a manner that IR2 occurs at the distal end of the REP2 ORF, without any intervening coding sequence, for example as seen in the 2μm plasmid. By "distal end" in this context we mean the end of the REP2 ORF opposite to the end from which the promoter initiates its transcription.

In one embodiment, the ORFs encoding REP2 and FLP may be present on the same region of the two regions defined between the inverted repeats of the 2μm-family plasmid, which region may be the bigger or smaller of the regions (if there is any inequality in size between the two regions).

In one embodiment, the ORFs encoding REP2 and FLP may be transcribed from divergent promoters.

Typically, the regions defined between the inverted repeats (e.g. such as defined as UL and US in the 2μm plasmid) of a 2μm-family plasmid may comprise not more than two endogenous genes that encode a protein that functions in the stable maintenance of the 2μm-family plasmid as a multicopy plasmid. Thus in a preferred embodiment, one region of the plasmid defined between the inverted repeats may comprise not more than the ORFs encoding FLP and REP2; FLP and REP1; or REP1 and REP2, as endogenous coding sequence.

A 2μm-family plasmid typically comprises an origin of replication (also known as an "autonomously replicating sequence - "ARS"), which is typically bidirectional. Any appropriate ARS sequence can be present. Consensus sequences typical of yeast chromosomal origins of replication may be appropriate (Broach *et al*, 1982, *Cold Spring Harbor Symp. Quant. Biol.*, **47**, 1165-1174; Williamson, *Yeast*, 1985, **1**, 1-14). Preferred ARSs include those isolated from pSR1, pSB1, pSB2, pSB3, pSB4, pSM1, pKD1, pPM1 and the 2μm plasmid.

Thus, a preferred 2μm-family plasmid may comprise ORFs encoding FLP, REP1 and REP2, two inverted repeat sequences each inverted repeat comprising an FRT site compatible with the encoded FLP protein, and an ARS sequence. Preferably the FRT sites are derived from the same 2μm-family plasmid as the sequence of the encoded FLP protein. More preferably the sequences of the encoded REP1 and REP2 proteins are derived from the same 2μm-family plasmid as each other. Even more preferably, the FRT sites are derived from the same 2μm-family plasmid as the sequence of the encoded FLP, REP1 and REP2 proteins. Yet more preferably, the sequences of the ORFs encoding FLP, REP1 and REP2, and the sequence of the inverted repeats (including the FRT sites) are derived

from the same 2µm-family plasmid. Furthermore, the ARS site may be derived from the same 2µm-family plasmid as one or more of the ORFs of FLP, REP1 and REP2, and the sequence of the inverted repeats (including the FRT sites).

The term "derived from" includes sequences having an identical sequence to the sequence from which they are derived. However, variants and fragments thereof, as defined above, are also included. For example, an *FLP* gene having a sequence derived from the *FLP* gene of the 2µm plasmid may have a modified promoter or other regulatory sequence compared to that of the naturally occurring gene. Additionally or alternatively, an *FLP* gene having a sequence derived from the *FLP* gene of the 2µm plasmid may have a modified nucleotide sequence in the open reading frame which may encode the same protein as the naturally occurring gene, or may encode a modified FLP protein. The same considerations apply to other sequences on a 2µm-family plasmid having a sequence derived from a particular source.

Optionally, a 2µm-family plasmid may comprise a region derived from the *STB* region (also known as REP3) of the 2µm plasmid, as defined in Volkert *et al*, *op. cit*. The *STB* region in a 2µm-family plasmid of the invention may comprise two or more tandem repeat sequences, such as three, four, five or more. Alternatively, no tandem repeat sequences may be present. The tandem repeats may be any size, such as 10, 20, 30, 40, 50, 60 70, 80, 90, 100 bp or more in length. The tandem repeats in the *STB* region of the 2µm plasmid are 62 bp in length. It is not essential for the sequences of the tandem repeats to be identical. Slight sequence variation can be tolerated. It may be preferable to select an *STB* region from the same

plasmid as either or both of the REP1 and REP2 ORFs. The *STB* region is thought to be a *cis*-acting element and preferably is not transcribed.

Optionally, a 2μm-family plasmid may comprise an additional ORF that encodes a protein that functions in the stable maintenance of the 2μm-family plasmid as a multicopy plasmid. The additional protein can be designated RAF or D. ORFs encoding the RAF or D gene can be seen on, for example, the 2μm plasmid and pSM1. Thus a RAF or D ORF can comprise a sequence suitable to encode the protein product of the RAF or D gene ORFs encoded by the 2μm plasmid or pSM1, or variants and fragments thereof. Thus variants and fragments of the protein products of the RAF or D genes of the 2μm plasmid or pSM1 are also included in the present invention. "Fragments" and "variants" of the protein products of the RAF or D genes of the 2μm plasmid or pSM1 are those which, when encoded by the 2μm plasmid or pSM1 in place of the native ORF, do not disrupt the stable multicopy maintenance of the plasmid within a suitable yeast population. Such variants and fragments will usually have at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, or more, homology with the protein product of the RAF or D gene ORFs encoded by the 2μm plasmid or pSM1.

A naturally occurring 2μm-family plasmid may be preferred. A naturally occurring 2μm-family plasmid is any plasmid having the features defined above, which plasmid is found to naturally exist in yeast, i.e. has not been recombinantly modified to include heterologous sequence. Preferably the naturally occurring 2μm-family plasmid is selected from pSR1 (Accession No. X02398), pSB3 (Accession No. X02608) or pSB4 as obtained from *Zygosaccharomyces rouxii*, pSB1 or pSB2 (Accession No. NC_002055 or

M18274) both as obtained from *Zygosaccharomyces bailli*, pSM1 (Accession No. NC_002054) as obtained from *Zygosaccharomyces fermentati*, pKD1 (Accession No. X03961) as obtained from *Kluyveromyces drosophilarum*, pPM1 from *Pichia membranaefaciens* or, most preferably, the 2μm plasmid (Accession No. NC_001398 or J01347) as obtained from *Saccharomyces cerevisiae*. Accession numbers in this paragraph refer to NCBI deposits.

The 2μm plasmid (Figure 1) is a 6,318-bp double-stranded DNA plasmid, endogenous in most *Saccharomyces cerevisiae* strains at 60-100 copies per haploid genome. The 2μm plasmid comprises a small unique (US) region and a large unique (UL) region, separated by two 599-bp inverted repeat sequences. Site-specific recombination of the inverted repeat sequences results in inter-conversion between the A-form and B-form of the plasmid *in vivo* (Volkert & Broach, 1986, *Cell*, **46**, 541). The two forms of 2μm differ only in the relative orientation of their unique regions.

While DNA sequencing of a cloned 2μm plasmid (also known as Scp1) from *Saccharomyces cerevisiae* gave a size of 6,318-bp (Hartley and Donelson, 1980, *Nature*, **286**, 860), other slightly smaller variants of 2μm, Scp2 and Scp3, are known to exist as a result of small deletions of 125-bp and 220-bp, respectively, in a region known as *STB* (Cameron *et al.*, 1977, *Nucl. Acids Res.*, **4**, 1429: Kikuchi, 1983, *Cell*, **35**, 487 and Livingston & Hahne, 1979, *Proc. Natl. Acad. Sci. USA*, **76**, 3727). In one study about 80% of natural *Saccharomyces* strains from around the world contained DNA homologous to 2μm (by Southern blot analysis) (Hollenberg, 1982, *Current Topics in Microbiology and Immunobiology*, **96**, 119). Furthermore, variation (genetic polymorphism) occurs within the natural population of 2μm plasmids found in *S. cerevisiae* and *S. carlsbergensis*,

with the NCBI sequence (accession number NC_001398) being one example.

The 2µm plasmid has a nuclear localisation and displays a high level of mitotic stability (Mead *et al*, 1986, *Molecular & General Genetics*, **205**, 417). The inherent stability of the 2µm plasmid results from a plasmid-encoded copy number amplification and partitioning mechanism, which can be compromised during the development of chimeric vectors (Futcher & Cox, 1984, *J. Bacteriol.*, **157**, 283; Bachmair & Ruis, 1984, *Monatshefte für Chemie*, **115**, 1229). A yeast strain, which contains a 2µm plasmid is known as [cir$^+$], while a yeast strain which does not contain a 2µm plasmid is known as [cir^0].

The US-region of the 2µm plasmid contains the *REP2* and *FLP* genes, and the UL-region contains the *REP1* and *D* (also known as *RAF*) genes, the *STB*-locus and the origin of replication (Broach & Hicks, 1980, *Cell*, **21**, 501; Sutton & Broach, 1985, *Mol. Cell. Biol.*, **5**, 2770). The Flp recombinase binds to FRT-sites (Flp Recognition Target) within the inverted repeats to mediate site-specific recombination, which is essential for natural plasmid amplification and control of plasmid copy number *in vivo* (Senecoff *et al*, 1985, *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 7270; Jayaram, 1985, *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 5875). The copy number of 2µm-family plasmids can be significantly affected by changes in Flp recombinase activity (Sleep *et al*, 2001, *Yeast*, **18**, 403; Rose & Broach, 1990, *Methods Enzymol.*, **185**, 234). The Rep1 and Rep2 proteins mediate plasmid segregation, although their mode of action is unclear (Sengupta *et al*, 2001, *J. Bacteriol.*, **183**, 2306). They also repress transcription of the *FLP* gene (Reynolds *et al*, 1987, *Mol. Cell. Biol.*, 7, 3566).

The *FLP* and *REP2* genes of the 2µm plasmid are transcribed from divergent promoters, with apparently no intervening sequence defined between them. The *FLP* and *REP2* transcripts both terminate at the same sequence motifs within the inverted repeat sequences, at 24-bp and 178-bp respectively after their translation termination codons (Sutton & Broach, 1985, *Mol. Cell. Biol.*, **5**, 2770).

In the case of *FLP*, the C-terminal coding sequence also lies within the inverted repeat sequence. Furthermore, the two inverted repeat sequences are highly conserved over 599-bp, a feature considered advantageous to efficient plasmid replication and amplification *in vivo*, although only the FRT-sites (less than 65-bp) are essential for site-specific recombination *in vitro* (Senecoff *et al*, 1985, *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 7270; Jayaram, 1985, *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 5875; Meyer-Leon *et al*, 1984, *Cold Spring Harbor Symposia On Quantitative Biology*, **49**, 797). The key catalytic residues of Flp are arginine-308 and tyrosine-343 (which is essential) with strand-cutting facilitated by histidine-309 and histidine 345 (Prasad *et al*, 1987, *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 2189; Chen *et al*, 1992, *Cell*, **69**, 647; Grainge *et al*, 2001, *J. Mol. Biol.*, **314**, 717).

Two functional domains are described in Rep2. Residues 15-58 form a Rep1-binding domain, and residues 59-296 contain a self-association and STB-binding region (Sengupta *et al*, 2001, *J. Bacteriol.*, **183**, 2306).

Chimeric or large deletion mutant derivatives of 2µm which lack many of the essential functional regions of the 2µm plasmid but retain the functional *cis* element *ARS* and *STB*, cannot effectively partition between mother and daughter cells at cell division. Such plasmids can do so if these functions

are supplied in *trans*, by for instance the provision of a functional 2µm plasmid within the host, such as a [cir$^+$] host.

Genes of interest have previously been inserted into the UL-region of the 2µm plasmid. For example, see plasmid pSAC3U1 in EP 0 286 424 and the plasmid shown in Figure 2, which includes a β-lactamase gene (for ampicillin resistance), a *LEU2* selectable marker and an oligonucleotide linker, the latter two of which are inserted into a unique *Sna*BI-site within the UL-region of the 2µm-like disintegration vector, pSAC3 (see EP 0 286 424). The *E. coli* DNA between the *Xba*I-sites that contains the ampicillin resistance gene is lost from the plasmid shown in Figure 2 after transformation into yeast. This is described in Chinery & Hinchliffe, 1989, *Curr. Genet.*, **16**, 21 and EP 0 286 424, where these types of vectors are designated "disintegration vectors". Further polynucleotide insertions can be made in a *Not*I-site within a linker (Sleep *et al*, 1991, *Biotechnology (N Y)*, 9, 183).

Alternative insertion sites in 2µm plasmid are known in the art, including those described in Rose & Broach (1990, *Methods Enzymol.*, **185**, 234-279), such as plasmids pCV19, pCV20, CV_{neo}, which utilise an insertion at *Eco*RI in *FLP*, plasmids pCV21, pGT41 and pYE which utilise *Eco*RI in *D* as the insertion site, plasmid pHKB52 which utilises *Pst*I in *D* as the insertion site, plasmid pJDB248 which utilises an insertion at *Pst*I in *D* and *Eco*RI in *D*, plasmid pJDB219 in which *Pst*I in *D* and *Eco*RI in *FLP* are used as insertion sites, plasmid G18, plasmid pAB18 which utilises an insertion at *Cla*I in *FLP*, plasmids pGT39 and pA3, plasmids pYT11, pYT14 and pYT11-LEU which use *Pst*I in *D* as the insertion site, and plasmid PTY39 which uses *Eco*RI in *FLP* as the insertion site. Other 2µm plasmids include pSAC3, pSAC3U1, pSAC3U2, pSAC300, pSAC310, pSAC3C1,

pSAC3PL1, pSAC3SL4, and pSAC3SC1 are described in EP 0 286 424 and Chinery & Hinchliffe (1989, *Curr. Genet.*, **16**, 21-25) which also described *Pst*I, *Eag*I or *Sna*BI as appropriate 2μm insertion sites. Further 2μm plasmids include pAYE255, pAYE316, pAYE443, pAYE522 (Kerry-Williams *et al*, 1998, *Yeast*, **14**, 161-169), pDB2244 (WO 00/44772), and pAYE329 (Sleep *et al*, 2001, *Yeast*, **18**, 403-421).

In one preferred embodiment, one or more genes are inserted into a 2μm-family plasmid within an untranscribed region around the ARS sequence. For example, in the 2μm plasmid obtained from *S. cerevisiae*, the untranscribed region around the ARS sequence extends from end of the D gene to the beginning of ARS sequence. Insertion into *Sna*BI (near the origin of replication sequence ARS) is described in Chinery & Hinchliffe, 1989, *Curr. Genet.*, **16**, 21-25. The skilled person will appreciate that gene insertions can also be made in the untranscribed region at neighbouring positions to the *Sna*BI site described in Chinery & Hinchliffe.

In another preferred embodiment, *REP2* and *FLP* genes in a 2μm-family plasmid each have an inverted repeat adjacent to them, and one or more genes are inserted into a 2μm-family plasmid within the region between the first base after the last functional codon of either the *REP2* gene or the *FLP* gene and the last base before the FRT site in the inverted repeat adjacent to said gene. The last functional codon of either a *REP2* gene or a *FLP* gene is the codon in the open reading frame of the gene that is furthest downstream from the promoter of the gene whose replacement by a stop codon will lead to an unacceptable loss of multicopy stability of the plasmid, as defined herein. Thus, disruption of the *REP2* or *FLP* genes at any point downstream of the last functional codon in either gene, by insertion of a

polynucleotide sequence insertion, deletion or substitution will not lead to an unacceptable loss of multicopy stability of the plasmid.

For example, the *REP2* gene of the 2µm plasmid can be disrupted after codon 59 and that the *FLP* gene of the 2µm plasmid can be disrupted after codon 344, each without a loss of multicopy stability of the plasmid. The last functional codon in equivalent genes in other 2µm-family plasmids can be determined routinely by making mutants of the plasmids in either the *FLP* or *REP2* genes and following the tests set out herein to determine whether the plasmid retains multicopy stability.

One can determined whether a plasmid retains multicopy stability using test such as defined in Chinery & Hinchliffe (1989, *Curr. Genet.*, **16**, 21-25). For yeast that do not grow in the non-selective media (YPD, also designated YEPD) defined in Chinery & Hinchliffe (1989, *Curr. Genet.*, **16**, 21-25) other appropriate non-selective media might be used. Plasmid stability may be defined as the percentage cells remaining prototrophic for the selectable marker after a defined number of generations. The number of generations will preferably be sufficient to show a difference between a control plasmid, such as pSAC35 or pSAC310, or to shown comparable stability to such a control plasmid. The number of generations may be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or more. Higher numbers are preferred. The acceptable plasmid stability might be 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.9% or substantially 100%. Higher percentages are preferred. The skilled person will appreciate that, even though a plasmid may have a stability less than 100% when grown on non-selective media, that plasmid can still be of use when cultured in selective media. For

example plasmid pDB2711 as described in the examples is only 10% stable when the stability is determined accordingly to test of Example 1, but provides a 15-fold increase in recombinant transferrin productivity in shake flask culture under selective growth conditions.

Thus one or more gene insertions may occur between the first base after the last functional codon of the *REP2* gene and the last base before the FRT site in an inverted repeat adjacent to said gene, more preferably between the first base of the inverted repeat and the last base before the FRT site, even more preferably at a position after the translation termination codon of the *REP2* gene and before the last base before the FRT site.

Additionally or alternatively one or more gene insertions may occur between the first base after the last functional codon of the *FLP* gene and the last base before the FRT site in an inverted repeat adjacent to said gene, preferably between the first base of the inverted repeat and the last base before the FRT site, more preferably between the first base after the end of the FLP coding sequence and the last base before the FRT site, such as at the first base after the end of the FLP coding sequence.

In one preferred embodiment, where the 2μm-family plasmid is based on the 2μm plasmid of *S. cerevisiae*, it is a disintegration vector as known in the art (for example, see EP 286 424, the contents of which are incorporated herein by reference). A disintegration vector may be a 2μm plasmid vector comprising a DNA sequence which is intended to be lost by recombination, three 2μm FRT sites, of which one pair of sites is in direct orientation and the other two pairs are in indirect orientation, and a DNA sequence of interest (such as an *E. coli* origin of replication and bacterial selectable

marker), the said sequence to be lost being located between the said sites which are in direct orientation.

Thus, the sequence to be lost may comprise a selectable marker DNA sequence.

A preferred disintegration vector comprises a complete 2μm plasmid additionally carrying (i) a bacterial plasmid DNA sequence necessary for propagation of the vector in a bacterial host; (ii) an extra 2μm FRT site; and a selectable marker DNA sequence for yeast transformation; the said bacterial plasmid DNA sequence being present and the extra FRT site being created at a restriction site, such as *Xba*I, in one of the two inverted repeat sequences of the 2μm plasmid, the said extra FRT site being in direct orientation in relation to the endogenous FRT site of the said one repeat sequence, and the bacterial plasmid DNA sequence being sandwiched between the extra FRT site and the endogenous FRT site of the said one repeat sequence. In a preferred disintegration vector, all bacterial plasmid DNA sequences are sandwiched as said. A particularly preferred 2μm plasmid vector has substantially the configuration of pSAC3 as shown in EP 286 424.

The term "disintegration vector" as used herein also includes plasmids as defined in US 6,451,559, the contents of which are incorporated herein by reference. Thus a disintegration vector may be a 2μm vector that, other than DNA sequence encoding non-yeast polypeptides, contains no bacterial (particularly *E. coli*) origin of replication, or more preferably no bacterial (particularly *E. coli*) sequence and preferably all DNA in said vector, other than DNA sequence encoding non-yeast polypeptides, is yeast-derived DNA.

The term "chaperone" as used herein refers to a protein that binds to and stabilises an otherwise unstable conformer of another protein, and by controlled binding and release, facilitates its correct fate *in vivo*, be it folding, oligomeric assembly, transport to a particular subcellular compartment, or disposal by degradation. Accordingly a chaperone is also a protein that is involved in protein folding, or which has chaperone activity or is involved in the unfolded protein response. Chaperone proteins of this type are known in the art, for example in the Stanford Genome Database (SGD), http:://db.yeastgenome.org. Preferred chaperones are eukaryotic chaperones, especially preferred chaperones are yeast chaperones, including *AHA1, CCT2, CCT3, CCT4, CCT5, CCT6, CCT7, CCT8, CNS1, CPR3, CPR6, ERO1, EUG1, FMO1, HCH1, HSP10, HSP12, HSP104, HSP26, HSP30, HSP42, HSP60, HSP78, HSP82, JEM1, MDJ1, MDJ2, MPD1, MPD2, PDI1, PFD1, ABC1, APJ1, ATP11, ATP12, BTT1, CDC37, CPR7, HSC82, KAR2, LHS1, MGE1, MRS11, NOB1, ECM10, SSA1, SSA2, SSA3, SSA4, SSC1, SSE2, SIL1, SLS1, ORM1, ORM2, PER1, PTC2, PSE1, UBI4* and *HAC1* or a truncated intronless *HAC1* (Valkonen *et al.* 2003, *Applied Environ. Micro.*, 69, 2065)

A preferred chaperone is protein disulphide isomerase (PDI) or a fragment or variant thereof having an equivalent ability to catalyse the formation of disulphide bonds within the lumen of the endoplasmic reticulum (ER). By "PDI" we include any protein having the ability to reactivate the ribonuclease activity against RNA of scrambled ribonuclease as described in EP 0 746 611 and Hillson *et al*, 1984, *Methods Enzymol.*, 107, 281-292.

PDI is an enzyme which typically catalyzes thiol:disulphide interchange reactions, and is a major resident protein component of the ER lumen in

secretory cells. A body of evidence suggests that it plays a role in secretory protein biosynthesis (Freedman, 1984, *Trends Biochem. Sci.*, **9**, 438-41) and this is supported by direct cross-linking studies *in situ* (Roth and Pierce, 1987, *Biochemistry*, **26**, 4179-82). The finding that microsomal membranes deficient in PDI show a specific defect in cotranslational protein disulphide (Bulleid and Freedman, 1988, *Nature*, **335**, 649-51) implies that the enzyme functions as a catalyst of native disulphide bond formation during the biosynthesis of secretory and cell surface proteins. This role is consistent with what is known of the enzyme's catalytic properties *in vitro*; it catalyzes thiol: disulphide interchange reactions leading to net protein disulphide formation, breakage or isomerization, and can typically catalyze protein folding and the formation of native disulphide bonds in a wide variety of reduced, unfolded protein substrates (Freedman *et al.*, 1989, *Biochem. Soc. Symp.*, **55**, 167-192). The DNA and amino acid sequence of the enzyme is known for several species (Scherens *et al*, 1991, *Yeast*, **7**, 185-193; Farquhar *et al*, 1991, *Gene*, **108**, 81-89; EP074661; EP0293793; EP0509841) and there is increasing information on the mechanism of action of the enzyme purified to homogeneity from mammalian liver (Creighton *et al*, 1980, *J. Mol. Biol.*, **142**, 43-62; Freedman *et al*, 1988, *Biochem. Soc. Trans.*, **16**, 96-9; Gilbert, 1989, *Biochemistry*, **28**, 7298-7305; Lundstrom and Holmgren, 1990, *J. Biol. Chem.*, **265**, 9114-9120; Hawkins and Freedman, 1990, *Biochem. J.*, **275**, 335-339). Of the many protein factors currently implicated as mediators of protein folding, assembly and translocation in the cell (Rothman, 1989, *Cell*, **59**, 591-601), PDI has a well-defined catalytic activity.

PDI is readily isolated from mammalian tissues and the homogeneous enzyme is a homodimer (2×57 kD) with characteristically acidic pI (4.0-4.5) (Hillson *et al*, 1984, *op. cit.*). The enzyme has also been purified from

wheat and from the alga *Chlamydomonas reinhardii* (Kaska *et al*, 1990, *Biochem. J.*, **268**, 63-68), rat (Edman *et al*, 1985, *Nature*, **317**, 267-270), bovine (Yamauchi *et al*, 1987, *Biochem. Biophys. Res. Comm.*, **146**, 1485-1492), human (Pihlajaniemi *et al*, 1987, *EMBO J.*, **6**, 643-9), yeast (Scherens *et al*, *supra*; Farquhar *et al*, *op. cit.*) and chick (Parkkonen *et al*, 1988, *Biochem. J.*, **256**, 1005-1011). The proteins from these vertebrate species show a high degree of sequence conservation throughout and all show several overall features first noted in the rat PDI sequence (Edman *et al.*, 1985, *op. cit.*).

Preferred PDI sequences include those from humans and those from yeast species, such as *S. cerevisiae*.

A yeast protein disulphide isomerase precursor, PDI1, can be found as Genbank accession no. CAA42373 or BAA00723. It has the following sequence of 522 amino acids:

```
1   mkfsagavls wssllllassv faqqeavape dsavvklatd sfneyiqshd lvlaeffapw
61  cghcknmape yvkaaetlve knitlaqidc tenqdlcmeh nipgfpslki fknsdvnnsi
121 dyegprtaea ivqfmikqsq pavavvadlp aylanetfvt pvivqsgkid adfnatfysm
181 ankhfndydf vsaenadddf klsiylpsam depvvyngkk adiadadvfe kwlqvealpy
241 fgeidgsvfa qyvesglplg ylfyndeeel eeykplftel akknrglmnf vsidarkfgr
301 hagnlnmkeq fplfaihdmt edlkyglpql seeafdelsd kivleskaie slvkdflkgd
361 aspivksqei fenqdssvfq lvgknhdeiv ndpkkdvlvl yyapwcghck rlaptyqela
421 dtyanatsdv liakldhten dvrgvviegy ptivlypggk ksesvvyqgs rsldslfdfi
481 kenghfdvdg kalyeeaqek aaeeadadae ladeedaihd el
```

An alternative yeast protein disulphide isomerase sequence can be found as Genbank accession no. CAA38402. It has the following sequence of 530 amino acids

```
1   mkfsagavls wsslllassv faqqeavape dsavvklatd sfneyiqshd lvlaeffapw
61  cghcknmape yvkaaetlve knitlaqidc tenqdlcmeh nipgfpslki fknrdvnnsi
121 dyegprtaea ivqfmikqsq pavavvadlp aylanetfvt pvivqsgkid adfnatfysm
181 ankhfndydf vsaenadddf klsiylpsam depvvyngkk adiadadvfe kwlqvealpy
241 fgeidgsvfa qyvesglplg ylfyndeeel eeykplftel akknrglmnf vsidarkfgr
301 hagnlnmkeq fplfaihdmt edlkyglpql seeafdelsd kivleskaie slvkdflkgd
361 aspivksqei fenqdssvfq lvgknhdeiv ndpkkdvlvl yyapwcghck rlaptyqela
421 dtyanatsdv liakldhten dvrgvviegy ptivlypggk ksesvvyqgs rsldslfdfi
481 kenghfdvdg kalyeeaqek aaeeaeadae aeadadaela deedaihdel
```

Variants and fragments of the above PDI sequences, and variants of other naturally occurring PDI sequences are also included in the present invention. A "variant", in the context of PDI, refers to a protein wherein at one or more positions there have been amino acid insertions, deletions, or substitutions, either conservative or non-conservative, provided that such changes result in a protein whose basic properties, for example enzymatic activity (type of and specific activity), thermostability, activity in a certain pH-range (pH-stability) have not significantly been changed. "Significantly" in this context means that one skilled in the art would say that the properties of the variant may still be different but would not be unobvious over the ones of the original protein.

By "conservative substitutions" is intended combinations such as Val, Ile, Leu, Ala, Met; Asp, Glu; Asn, Gln; Ser, Thr, Gly, Ala; Lys, Arg, His; and Phe, Tyr, Trp. Preferred conservative substitutions include Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr.

A "variant" typically has at least 25%, at least 50%, at least 60% or at least 70%, preferably at least 80%, more preferably at least 90%, even more preferably at least 95%, yet more preferably at least 99%, most preferably at least 99.5% sequence identity to the polypeptide from which it is derived.

The percent sequence identity between two polypeptides may be determined using suitable computer programs, as discussed below. Such variants may be natural or made using the methods of protein engineering and site-directed mutagenesis as are well known in the art.

A "fragment", in the context of PDI, refers to a protein wherein at one or more positions there have been deletions. Thus the fragment may comprise at most 5, 10, 20, 30, 40 or 50% of the complete sequence of the full mature PDI protein. Typically a fragment comprises up to 60%, more typically up to 70%, preferably up to 80%, more preferably up to 90%, even more preferably up to 95%, yet more preferably up to 99% of the complete sequence of the full PDI protein. Particularly preferred fragments of PDI protein comprise one or more whole domains of the desired protein.

A gene encoding a protein comprising the sequence of a chaperone may be formed in a like manner to that discussed below for genes encoding heterologous proteins, with particular emphasis on combinations of ORFs and regulatory regions.

The term "protein" as used herein includes all natural and non-natural proteins, polypeptides and peptides. A "heterologous protein" is a protein that is not naturally encoded by a 2μm-family plasmid. Preferably, therefore, the heterologous protein is not a FLP, REP1, REP2, or a RAF/D protein as encoded by any one of pSR1, pSB3 or pSB4 as obtained from *Z. rouxii*, pSB1 or pSB2 both as obtained from *Z. bailli*, pSM1 as obtained from *Z. fermentati*, pKD1 as obtained from *K. drosophilarum*, pPM1 as obtained from *P. membranaefaciens* or the 2μm plasmid as obtained from *S. cerevisiae*.

A gene encoding a heterologous protein comprises polynucleotide sequence encoding the heterologous protein (typically according to standard codon usage for any given organism), designated the open reading frame ("ORF"). The gene may additionally comprise some polynucleotide sequence that does not encode an open reading frame (termed "non-coding region").

Non-coding region in the gene may contain one or more regulatory sequences, operatively linked to the ORF, which allow for the transcription of the open reading frame and/or translation of the resultant transcript.

The term "regulatory sequence" refers to a sequence that modulates (i.e., promotes or reduces) the expression (i.e., the transcription and/or translation) of an ORF to which it is operably linked. Regulatory regions typically include promoters, terminators, ribosome binding sites and the like. The skilled person will appreciate that the choice of regulatory region will depend upon the intended expression system. For example, promoters may be constitutive or inducible and may be cell- or tissue-type specific or non-specific.

Suitable regulatory regions, may be 5bp, 10bp, 15bp, 20bp, 25bp, 30bp, 35bp, 40bp, 45bp, 50bp, 60bp, 70bp, 80bp, 90bp, 100bp, 120bp, 140bp, 160bp, 180bp, 200bp, 220bp, 240bp, 260bp, 280bp, 300bp, 350bp, 400bp, 450bp, 500bp, 550bp, 600bp, 650bp, 700bp, 750bp, 800bp, 850bp, 900bp, 950bp, 1000bp, 1100bp, 1200bp, 1300bp, 1400bp, 1500bp or greater, in length.

Those skilled in the art will recognise that the gene encoding the chaperone, for example PDI, may additionally comprise non-coding regions and/or regulatory regions. Such non-coding regions and regulatory regions are not

restricted to the native non-coding regions and/or regulatory regions normally associated with the chaperone ORF.

Where the expression system is yeast, such as *Saccharomyces cerevisiae*, suitable promoters for *S. cerevisiae* include those associated with the *PGK1* gene, *GAL1* or *GAL10* genes, *TEF1*, *TEF2*, *PYK1*, *PMA1*, *CYC1*, *PHO5*, *TRP1*, *ADH1*, *ADH2*, the genes for glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, triose phosphate isomerase, phosphoglucose isomerase, glucokinase, α-mating factor pheromone, a-mating factor pheromone, the *PRB1* promoter, the *PRA1* promoter, the *GPD1* promoter, and hybrid promoters involving hybrids of parts of 5' regulatory regions with parts of 5' regulatory regions of other promoters or with upstream activation sites (e.g. the promoter of EP-A-258 067).

Suitable transcription termination signals are well known in the art. Where the host cell is eukaryotic, the transcription termination signal is preferably derived from the 3' flanking sequence of a eukaryotic gene, which contains proper signals for transcription termination and polyadenylation. Suitable 3' flanking sequences may, for example, be those of the gene naturally linked to the expression control sequence used, i.e. may correspond to the promoter. Alternatively, they may be different. In that case, and where the host is a yeast, preferably *S. cerevisiae*, then the termination signal of the *S. cerevisiae ADH1*, *ADH2*, *CYC1*, or *PGK1* genes are preferred.

In one embodiment, the favoured regulatory sequences in yeast, such as *Saccharomyces cerevisiae*, include: a yeast promoter (e.g. the *Saccharomyces cerevisiae PRB1* promoter), as taught in EP 431 880; and a transcription terminator, preferably the terminator from *Saccharomyces*

ADH1, as taught in EP 60 057. Preferably, the vector incorporates at least two translation stop codons.

It may be beneficial for the non-coding region to incorporate more than one DNA sequence encoding a translational stop codon, such as UAA, UAG or UGA, in order to minimise translational read-through and thus avoid the production of elongated, non-natural fusion proteins. The translation stop codon UAA is preferred.

The term "operably linked" includes within its meaning that a regulatory sequence is positioned within any non-coding region in a gene such that it forms a relationship with an ORF that permits the regulatory region to exert an effect on the ORF in its intended manner. Thus a regulatory region "operably linked" to an ORF is positioned in such a way that the regulatory region is able to influence transcription and/or translation of the ORF in the intended manner, under conditions compatible with the regulatory sequence.

In one preferred embodiment, the heterologous protein is secreted. In that case, a sequence encoding a secretion leader sequence which, for example, comprises most of the natural HSA secretion leader, plus a small portion of the *S. cerevisiae* α-mating factor secretion leader as taught in WO 90/01063 may be included in the open reading frame.

Alternatively, the heterologous protein may be intracellular.

In another preferred embodiment, the heterologous protein comprises the sequence of a eukaryotic protein, or a fragment or variant thereof. Suitable eukaryotes include fungi, plants and animals. In one preferred embodiment

the heterologous protein is a fungal protein, such as a yeast protein. In another preferred embodiment the heterologous protein is an animal protein. Exemplary animals include vertebrates and invertebrates. Exemplary vertebrates include mammals, such as humans, and non-human mammals.

Thus the heterologous protein may comprise the sequence of a yeast protein. It may, for example, comprise the sequence of a yeast protein from the same host from which the 2μm-family plasmid is derived. Those skilled in the art will recognise that a method, use or plasmid of the first, second or third aspects of the invention may comprise DNA sequences encoding more than one heterologous protein, more than one chaperone, or more than one heterologous protein and more than one chaperone.

In another preferred embodiment, the heterologous protein may comprise the sequence of albumin, a monoclonal antibody, an etoposide, a serum protein (such as a blood clotting factor), antistasin, a tick anticoagulant peptide, transferrin, lactoferrin, endostatin, angiostatin, collagens, immunoglobulins, Fab' fragments, F(ab')2, scAb, scFv, interferons, interleukins, IL10, IL11, IL2, interferon α species and sub-species, interferon β species and sub-species, interferon γ species and sub-species, leptin, CNTF, $CNTF_{Ax15}$, IL1-receptor antagonist, erythropoietin (EPO) and EPO mimics, thrombopoietin (TPO) and TPO mimics, prosaptide, cyanovirin-N, 5-helix, T20 peptide, T1249 peptide, HIV gp41, HIV gp120, urokinase, prourokinase, tPA, hirudin, platelet derived growth factor, parathyroid hormone, proinsulin, insulin, glucagon, glucagon-like peptides, insulin-like growth factor, calcitonin, growth hormone, transforming growth factor β, tumour necrosis factor, G-CSF, GM-CSF, M-CSF, FGF, coagulation factors in both pre and active forms, including but not limited to plasminogen, fibrinogen, thrombin, pre-thrombin, pro-thrombin, von

Willebrand's factor, α_1-antitrypsin, plasminogen activators, Factor VII, Factor VIII, Factor IX, Factor X and Factor XIII, nerve growth factor, LACI, platelet-derived endothelial cell growth factor (PD-ECGF), glucose oxidase, serum cholinesterase, aprotinin, amyloid precursor protein, inter-alpha trypsin inhibitor, antithrombin III, apo-lipoprotein species, Protein C, Protein S, or a variant or fragment of any of the above.

A "variant", in the context of the above-listed proteins, refers to a protein wherein at one or more positions there have been amino acid insertions, deletions, or substitutions, either conservative or non-conservative, provided that such changes result in a protein whose basic properties, for example enzymatic activity or receptor binding (type of and specific activity), thermostability, activity in a certain pH-range (pH-stability) have not significantly been changed. "Significantly" in this context means that one skilled in the art would say that the properties of the variant may still be different but would not be unobvious over the ones of the original protein.

By "conservative substitutions" is intended combinations such as Val, Ile, Leu, Ala, Met; Asp, Glu; Asn, Gln; Ser, Thr, Gly, Ala; Lys, Arg, His; and Phe, Tyr, Trp. Preferred conservative substitutions include Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr.

A "variant" typically has at least 25%, at least 50%, at least 60% or at least 70%, preferably at least 80%, more preferably at least 90%, even more preferably at least 95%, yet more preferably at least 99%, most preferably at least 99.5% sequence identity to the polypeptide from which it is derived.

The percent sequence identity between two polypeptides may be determined using suitable computer programs, for example the GAP program of the

University of Wisconsin Genetic Computing Group and it will be appreciated that percent identity is calculated in relation to polypeptides whose sequence has been aligned optimally.

The alignment may alternatively be carried out using the Clustal W program (Thompson *et al.*, (1994) *Nucleic Acids Res.*, 22(22), 4673-80). The parameters used may be as follows:

- Fast pairwise alignment parameters: K-tuple(word) size; 1, window size; 5, gap penalty; 3, number of top diagonals; 5. Scoring method: x percent.
- Multiple alignment parameters: gap open penalty; 10, gap extension penalty; 0.05.
- Scoring matrix: BLOSUM.

Such variants may be natural or made using the methods of protein engineering and site-directed mutagenesis as are well known in the art.

A "fragment", in the context of the above-listed proteins, refers to a protein wherein at one or more positions there have been deletions. Thus the fragment may comprise at most 5, 10, 20, 30, 40 or 50% of the complete sequence of the full mature polypeptide. Typically a fragment comprises up to 60%, more typically up to 70%, preferably up to 80%, more preferably up to 90%, even more preferably up to 95%, yet more preferably up to 99% of the complete sequence of the full desired protein. Particularly preferred fragments of a protein comprise one or more whole domains of the protein.

In one particularly preferred embodiment the heterologous protein comprises the sequence of albumin or a variant or fragment thereof.

By "albumin" we include a protein comprising the sequence of an albumin protein obtained from any source. Typically the source is mammalian. In one preferred embodiment the serum albumin is human serum albumin ("HSA"). The term "human serum albumin" includes the meaning of a serum albumin having an amino acid sequence naturally occurring in humans, and variants thereof. Preferably the albumin has the amino acid sequence disclosed in WO 90/13653 or a variant thereof. The HSA coding sequence is obtainable by known methods for isolating cDNA corresponding to human genes, and is also disclosed in, for example, EP 73 646 and EP 286 424.

In another preferred embodiment the "albumin" comprises the sequence of bovine serum albumin. The term "bovine serum albumin" includes the meaning of a serum albumin having an amino acid sequence naturally occurring in cows, for example as taken from Swissprot accession number P02769, and variants thereof as defined below. The term "bovine serum albumin" also includes the meaning of fragments of full-length bovine serum albumin or variants thereof, as defined below.

In another preferred embodiment the albumin comprises the sequence of an albumin derived from one of serum albumin from dog (e.g. see Swissprot accession number P49822), pig (e.g. see Swissprot accession number P08835), goat (e.g. as available from Sigma as product no. A2514 or A4164), turkey (e.g. see Swissprot accession number O73860), baboon (e.g. as available from Sigma as product no. A1516), cat (e.g. see Swissprot accession number P49064), chicken (e.g. see Swissprot accession number P19121), ovalbumin (e.g. chicken ovalbumin) (e.g. see Swissprot accession number P01012), donkey (e.g. see Swissprot accession number P39090), guinea pig (e.g. as available from Sigma as product no. A3060, A2639,

O5483 or A6539), hamster (e.g. as available from Sigma as product no. A5409), horse (e.g. see Swissprot accession number P35747), rhesus monkey (e.g. see Swissprot accession number Q28522), mouse (e.g. see Swissprot accession number O89020), pigeon (e.g. as defined by Khan *et al*, 2002, *Int. J. Biol. Macromol.*, 30(3-4),171-8), rabbit (e.g. see Swissprot accession number P49065), rat (e.g. see Swissprot accession number P36953) and sheep (e.g. see Swissprot accession number P14639) and includes variants and fragments thereof as defined below.

Many naturally occurring mutant forms of albumin are known. Many are described in Peters, (1996, *All About Albumin: Biochemistry, Genetics and Medical Applications*, Academic Press, Inc., San Diego, California, p.170-181). A variant as defined above may be one of these naturally occurring mutants.

A "variant albumin" refers to an albumin protein wherein at one or more positions there have been amino acid insertions, deletions, or substitutions, either conservative or non-conservative, provided that such changes result in an albumin protein for which at least one basic property, for example binding activity (type of and specific activity e.g. binding to bilirubin), osmolarity (oncotic pressure, colloid osmotic pressure), behaviour in a certain pH-range (pH-stability) has not significantly been changed. "Significantly" in this context means that one skilled in the art would say that the properties of the variant may still be different but would not be unobvious over the ones of the original protein.

By "conservative substitutions" is intended combinations such as Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr. Such variants may be made by techniques well known in the art, such as by site-

directed mutagenesis as disclosed in US Patent No 4,302,386 issued 24 November 1981 to Stevens, incorporated herein by reference.

Typically an albumin variant will have more than 40%, usually at least 50%, more typically at least 60%, preferably at least 70%, more preferably at least 80%, yet more preferably at least 90%, even more preferably at least 95%, most preferably at least 98% or more sequence identity with naturally occurring albumin. The percent sequence identity between two polypeptides may be determined using suitable computer programs, for example the GAP program of the University of Wisconsin Genetic Computing Group and it will be appreciated that percent identity is calculated in relation to polypeptides whose sequence has been aligned optimally. The alignment may alternatively be carried out using the Clustal W program (Thompson *et al.*, 1994). The parameters used may be as follows:

Fast pairwise alignment parameters: K-tuple(word) size; 1, window size; 5, gap penalty; 3, number of top diagonals; 5. Scoring method: x percent. Multiple alignment parameters: gap open penalty; 10, gap extension penalty; 0.05. Scoring matrix: BLOSUM.

The term "fragment" as used above includes any fragment of full-length albumin or a variant thereof, so long as at least one basic property, for example binding activity (type of and specific activity e.g. binding to bilirubin), osmolarity (oncotic pressure, colloid osmotic pressure), behaviour in a certain pH-range (pH-stability) has not significantly been changed. "Significantly" in this context means that one skilled in the art would say that the properties of the variant may still be different but would not be unobvious over the ones of the original protein. A fragment will typically be at least 50 amino acids long. A fragment may comprise at least one whole sub-domain

of albumin. Domains of HSA have been expressed as recombinant proteins (Dockal, M. *et al.*, 1999, *J. Biol. Chem.*, **274**, 29303-29310), where domain I was defined as consisting of amino acids 1-197, domain II was defined as consisting of amino acids 189-385 and domain III was defined as consisting of amino acids 381-585. Partial overlap of the domains occurs because of the extended α-helix structure (h10-h1) which exists between domains I and II, and between domains II and III (Peters, 1996, *op. cit.*, Table 2-4). HSA also comprises six sub-domains (sub-domains IA, IB, IIA, IIB, IIIA and IIIB). Sub-domain IA comprises amino acids 6-105, sub-domain IB comprises amino acids 120-177, sub-domain IIA comprises amino acids 200-291, sub-domain IIB comprises amino acids 316-369, sub-domain IIIA comprises amino acids 392-491 and sub-domain IIIB comprises amino acids 512-583. A fragment may comprise a whole or part of one or more domains or sub-domains as defined above, or any combination of those domains and/or sub-domains.

In another particularly preferred embodiment the heterologous protein comprises the sequence of transferrin or a variant or fragment thereof. The term "transferrin" as used herein includes all members of the transferrin family (Testa, *Proteins of iron metabolism*, CRC Press, 2002; Harris & Aisen, *Iron carriers and iron proteins*, Vol. 5, Physical Bioinorganic Chemistry, VCH, 1991) and their derivatives, such as transferrin, mutant transferrins (Mason *et al*, 1993, *Biochemistry*, **32**, 5472; Mason *et al*, 1998, *Biochem. J.*, **330**(1), 35), truncated transferrins, transferrin lobes (Mason *et al*, 1996, *Protein Expr. Purif.*, **8**, 119; Mason *et al*, 1991, *Protein Expr. Purif.*, **2**, 214), lactoferrin, mutant lactoferrins, truncated lactoferrins, lactoferrin lobes or fusions of any of the above to other peptides, polypeptides or proteins (Shin *et al*, 1995, *Proc. Natl. Acad. Sci. USA*, **92**,

2820; Ali *et al*, 1999, *J. Biol. Chem.*, **274**, 24066; Mason *et al*, 2002, *Biochemistry*, **41**, 9448).

The transferrin may be human transferrin. The term "human transferrin" is used herein to denote material which is indistinguishable from transferrin derived from a human or which is a variant or fragment thereof. A "variant" includes insertions, deletions and substitutions, either conservative or non-conservative, where such changes do not substantially alter the useful ligand-binding or immunogenic properties of transferrin.

Mutants of transferrin are included in the invention. Such mutants may have altered immunognenicity. For example, transferrin mutants may display modified (e.g. reduced) glycosylation. The N-linked glycosylation pattern of a transferrin molecule can be modified by adding/removing amino acid glycosylation consensus sequences such as N-X-S/T, at any or all of the N, X, or S/T position. An example of a transferrin mutant modified in this manner is exemplified below.

We also include naturally-occurring polymorphic variants of human transferrin or human transferrin analogues. Generally, variants or fragments of human transferrin will have at least 50% (preferably at least 80%, 90% or 95%) of human transferrin's ligand binding activity (for example iron-binding), weight for weight. The iron binding activity of transferrin or a test sample can be determined spectrophotometrically by 470nm:280nm absorbance ratios for the proteins in their iron-free and fully iron-loaded states. Reagents should be iron-free unless stated otherwise. Iron can be removed from transferrin or the test sample by dialysis against 0.1M citrate, 0.1M acetate, 10mM EDTA pH4.5. Protein should be at approximately 20mg/mL in 100mM HEPES, 10mM $NaHCO_3$ pH8.0. Measure the

470nm:280nm absorbance ratio of apo-transferrin (Calbiochem, CN Biosciences, Nottingham, UK) diluted in water so that absorbance at 280nm can be accurately determined spectrophotometrically (0% iron binding). Prepare 20mM iron-nitrilotriacetate (FeNTA) solution by dissolving 191mg nitrotriacetic acid in 2mL 1M NaOH, then add 2mL 0.5M ferric chloride. Dilute to 50mL with deionised water. Fully load apo-transferrin with iron (100% iron binding) by adding a sufficient excess of freshly prepared 20mM FeNTA, then dialyse the holo-transferrin preparation completely against 100mM HEPES, 10mM $NaHCO_3$ pH8.0 to remove remaining FeNTA before measuring the absorbance ratio at 470nm:280nm. Repeat the procedure using test sample, which should initially be free from iron, and compare final ratios to the control.

Additionally, single or multiple heterologous fusions comprising any of the above; or single or multiple heterologous fusions to albumin, transferrin or immunoglobins or a variant or fragment of any of these may be used. Such fusions include albumin N-terminal fusions, albumin C-terminal fusions and co-N-terminal and C-terminal albumin fusions as exemplified by WO 01/79271, and transferrin N-terminal fusions, transferrin C-terminal fusions, and co-N-terminal and C-terminal transferrin fusions.

Examples of transferrin fusions are given in US patent applications US2003/0221201 and US2003/0226155, Shin, *et al.*, 1995, *Proc Natl Acad Sci U S A*, **92**, 2820, Ali, *et al.*, 1999, *J Biol Chem*, **274**, 24066, Mason, *et al.*, 2002, *Biochemistry*, **41**, 9448, the contents of which are incorporated herein by reference.

The skilled person will also appreciate that the open reading frame of any other gene or variant, or part or either, can be utilised as an open reading

frame for use with the present invention. For example, the open reading frame may encode a protein comprising any sequence, be it a natural protein (including a zymogen), or a variant, or a fragment (which may, for example, be a domain) of a natural protein; or a totally synthetic protein; or a single or multiple fusion of different proteins (natural or synthetic). Such proteins can be taken, but not exclusively, from the lists provided in WO 01/79258, WO 01/79271, WO 01/79442, WO 01/79443, WO 01/79444 and WO 01/79480, or a variant or fragment thereof; the disclosures of which are incorporated herein by reference. Although these patent applications present the list of proteins in the context of fusion partners for albumin, the present invention is not so limited and, for the purposes of the present invention, any of the proteins listed therein may be presented alone or as fusion partners for albumin, the Fc region of immunoglobulin, transferrin, lactoferrin or any other protein or fragment or variant of any of the above, as a desired polypeptide.

The heterologous protein may be a therapeutically active protein. In other words, it may have a recognised medical effect on individuals, such as humans. Many different types of therapeutically active protein are well known in the art.

The heterologous protein may comprise a leader sequence effective to cause secretion in yeast.

Numerous natural or artificial polypeptide signal sequences (also called secretion pre regions) have been used or developed for secreting proteins from host cells. The signal sequence directs the nascent protein towards the machinery of the cell that exports proteins from the cell into the surrounding medium or, in some cases, into the periplasmic space. The

signal sequence is usually, although not necessarily, located at the N-terminus of the primary translation product and is generally, although not necessarily, cleaved off the protein during the secretion process, to yield the "mature" protein.

In the case of some proteins the entity that is initially secreted, after the removal of the signal sequence, includes additional amino acids at its N-terminus called a "pro" sequence, the intermediate entity being called a "pro-protein". These pro sequences may assist the final protein to fold and become functional, and are usually then cleaved off. In other instances, the pro region simply provides a cleavage site for an enzyme to cleave off the pre-pro region and is not known to have another function.

The pro sequence can be removed either during the secretion of the protein from the cell or after export from the cell into the surrounding medium or periplasmic space.

Polypeptide sequences which direct the secretion of proteins, whether they resemble signal (i.e. pre) sequences or pre-pro secretion sequences, are referred to as leader sequences. The secretion of proteins is a dynamic process involving translation, translocation and post-translational processing, and one or more of these steps may not necessarily be completed before another is either initiated or completed.

For production of proteins in eukaryotic species such as the yeasts *Saccharomyces cerevisiae*, *Zygosaccharomyces* species, *Kluyveromyces lactis* and *Pichia pastoris*, known leader sequences include those from the *S. cerevisiae* acid phosphatase protein (Pho5p) (see EP 366 400), the invertase protein (Suc2p) (see Smith *et al.* (1985) *Science*, 229, 1219-1224)

and heat-shock protein-150 (Hsp150p) (see WO 95/33833). Additionally, leader sequences from the *S. cerevisiae* mating factor alpha-1 protein (MFα-1) and from the human lysozyme and human serum albumin (HSA) protein have been used, the latter having been used especially, although not exclusively, for secreting human albumin. WO 90/01063 discloses a fusion of the MFα-1 and HSA leader sequences, which advantageously reduces the production of a contaminating fragment of human albumin relative to the use of the MFα-1 leader sequence. Modified leader sequences are also disclosed in the examples of this application and the reader will appreciate that those leader sequences can be used with proteins other than transferrin.

Where the chaperone is protein disulphide isomerase, then preferably the heterologous protein comprises disulphide bonds in its mature form. The disulphide bonds may be intramolecular and/or intermolecular.

The heterologous protein may be a commercially useful protein. Some heterologously expressed proteins are intended to interact with the cell in which they are expressed in order to bring about a beneficial effect on the cell's activities. These proteins are not, in their own right, commercially useful. Commercially useful proteins are proteins that have a utility *ex vivo* of the cell in which they are expressed. Nevertheless, the skilled reader will appreciate that a commercially useful protein may also have a biological effect on the host cell expressing it as a heterologous protein, but that that effect is not the main or sole reason for expressing the protein therein.

In one embodiment it is preferred that the heterologous protein is not β-lactamase. In another embodiment it is preferred that the heterologous protein is not antistasin. However, the reader will appreciate that neither of these provisos exclude genes encoding either β-lactamase or antistasin from

being present on the 2μm-family plasmid of the invention, merely that the gene encoding the heterologous protein encodes a protein other than β-lactamase and/or antistasin.

Plasmids can be prepared by modifying 2μm-family plasmids known in the art by inserting a gene encoding a chaperone and inserting a gene encoding a heterologous protein using techniques well known in the art such as are described in by Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2001, 3rd edition, the contents of which are incorporated herein by reference. For example, one such method involves ligation via cohesive ends. Compatible cohesive ends can be generated on a DNA fragment for insertion and plasmid by the action of suitable restriction enzymes. These ends will rapidly anneal through complementary base pairing and remaining nicks can be closed by the action of DNA ligase.

A further method uses synthetic double stranded oligonucleotide linkers and adaptors. DNA fragments with blunt ends are generated by bacteriophage T4 DNA polymerase or *E.coli* DNA polymerase I which remove protruding 3' termini and fill in recessed 3' ends. Synthetic linkers and pieces of blunt-ended double-stranded DNA, which contain recognition sequences for defined restriction enzymes, can be ligated to blunt-ended DNA fragments by T4 DNA ligase. They are subsequently digested with appropriate restriction enzymes to create cohesive ends and ligated to an expression vector with compatible termini. Adaptors are also chemically synthesised DNA fragments which contain one blunt end used for ligation but which also possess one preformed cohesive end. Alternatively a DNA fragment or DNA fragments can be ligated together by the action of DNA ligase in the presence or absence of one or more synthetic double stranded oligonucleotides optionally containing cohesive ends.

Synthetic linkers containing a variety of restriction endonuclease sites are commercially available from a number of sources including Sigma-Genosys Ltd, London Road, Pampisford, Cambridge, United Kingdom.

Appropriate insertion sites in 2µm-family plasmids include, but are not limited to, those discussed above.

The present invention also provides a host cell comprising a plasmid as defined above. The host cell may be any type of cell. Bacterial and yeast host cells are preferred. Bacterial host cells may be useful for cloning purposes. Yeast host cells may be useful for expression of genes present in the plasmid.

In one embodiment the host cell is a yeast cell, such as a member of the *Saccharomyces*, *Kluyveromyces*, or *Pichia* genus, such *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, *Pichia pastoris and Pichia membranaefaciens, or Zygosaccharomyces rouxii*, *Zygosaccharomyces bailii*, *Zygosaccharomyces fermentati*, or *Kluyveromyces drosphilarum* are preferred.

The host cell type may be selected for compatibility with the plasmid type being used. Plasmids obtained from one yeast type can be maintained in other yeast types (Irie *et al*, 1991, *Gene*, **108**(1), 139-144; Irie *et al*, 1991, *Mol. Gen. Genet.*, **225**(2), 257-265). For example, pSR1 from *Zygosaccharomyces rouxii* can be maintained in *Saccharomyces cerevisiae*. Preferably, the host cell is compatible with the 2µm-family plasmid used (see below for a full description of the following plasmids). For example, where the plasmid is based on pSR1, pSB3 or pSB4 then a suitable yeast

cell is *Zygosaccharomyces rouxii*; where the plasmid is based on pSB1 or pSB2 then a suitable yeast cell is *Zygosaccharomyces bailli*; where the plasmid is based on pSM1 then a suitable yeast cell is *Zygosaccharomyces fermentati*; where the plasmid is based on pKD1 then a suitable yeast cell is *Kluyveromyces drosophilarum*; where the plasmid is based on pPM1 then a suitable yeast cell is *Pichia membranaefaciens*; where the plasmid is based on the 2μm plasmid then a suitable yeast cell is *Saccharomyces cerevisiae* or *Saccharomyces carlsbergensis*. It is particularly preferred that the plasmid is based on the 2μm plasmid and the yeast cell is *Saccharomyces cerevisiae*.

A 2μm-family plasmid of the invention can be said to be "based on" a naturally occurring plasmid if it comprises one, two or preferably three of the genes *FLP*, *REP1* and *REP2* having sequences derived from that naturally occurring plasmid.

It may be particularly advantageous to use a yeast deficient in one or more protein mannosyl transferases involved in O-glycosylation of proteins, for instance by disruption of the gene coding sequence.

Recombinantly expressed proteins can be subject to undesirable post-translational modifications by the producing host cell. For example, the albumin protein sequence does not contain any sites for N-linked glycosylation and has not been reported to be modified, in nature, by O-linked glycosylation. However, it has been found that recombinant human albumin ("rHA") produced in a number of yeast species can be modified by O-linked glycosylation, generally involving mannose. The mannosylated albumin is able to bind to the lectin Concanavalin A. The amount of mannosylated albumin produced by the yeast can be reduced by using a

yeast strain deficient in one or more of the *PMT* genes (WO 94/04687). The most convenient way of achieving this is to create a yeast which has a defect in its genome such that a reduced level of one of the Pmt proteins is produced. For example, there may be a deletion, insertion or transposition in the coding sequence or the regulatory regions (or in another gene regulating the expression of one of the *PMT* genes) such that little or no Pmt protein is produced. Alternatively, the yeast could be transformed to produce an anti-Pmt agent, such as an anti-Pmt antibody.

If a yeast other than *S. cerevisiae* is used, disruption of one or more of the genes equivalent to the *PMT* genes of *S. cerevisiae* is also beneficial, e.g. in *Pichia pastoris* or *Kluyveromyces lactis*. The sequence of *PMT1* (or any other *PMT* gene) isolated from *S. cerevisiae* may be used for the identification or disruption of genes encoding similar enzymatic activities in other fungal species. The cloning of the *PMT1* homologue of *Kluyveromyces lactis* is described in WO 94/04687.

The yeast will advantageously have a deletion of the *HSP150* and/or *YAP3* genes as taught respectively in WO 95/33833 and WO 95/23857.

A plasmid as defined above, may be introduced into a host through standard techniques. With regard to transformation of prokaryotic host cells, see, for example, Cohen *et al* (1972) *Proc. Natl. Acad. Sci. USA* **69**, 2110 and Sambrook *et al* (2001) *Molecular Cloning, A Laboratory Manual*, 3rd Ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. Transformation of yeast cells is described in Sherman *et al* (1986) *Methods In Yeast Genetics, A Laboratory Manual*, Cold Spring Harbor, NY. The method of Beggs (1978) *Nature* 275, 104-109 is also useful. Methods for the transformation of *S. cerevisiae* are taught generally in EP 251 744, EP 258 067 and WO

90/01063, all of which are incorporated herein by reference. With regard to vertebrate cells, reagents useful in transfecting such cells, for example calcium phosphate and DEAE-dextran or liposome formulations, are available from Stratagene Cloning Systems, or Life Technologies Inc., Gaithersburg, MD 20877, USA.

Electroporation is also useful for transforming cells and is well known in the art for transforming yeast cell, bacterial cells and vertebrate cells. Methods for transformation of yeast by electroporation are disclosed in Becker & Guarente (1990) *Methods Enzymol.* **194**, 182.

Generally, the plasmid will transform not all of the hosts and it will therefore be necessary to select for transformed host cells. Thus, a plasmid may comprise a selectable marker, including but not limited to bacterial selectable marker and/or a yeast selectable marker. A typical bacterial selectable marker is the β-lactamase gene although many others are known in the art. Typical yeast selectable marker include *LEU2, TRP1, HIS3, HIS4, URA3, URA5, SFA1, ADE2, MET15, LYS5, LYS2, ILV2, FBA1* and *PGK1*. Those skilled in the art will appreciate that any gene whose chromosomal deletion or inactivation results in an inviable host, so called essential genes, can be used as a selective marker if a functional gene is provided on the plasmid, as demonstrated for *PGK1* in a *pgk1* yeast strain (Piper and Curran, 1990, *Curr. Genet.* **17**, 119). Suitable essential genes can be found within the Stanford Genome Database (SGD), http:://db.yeastgenome.org).

Additionally, a plasmid according to any one of the first, second or third aspects of the present invention may comprise more than one selectable marker.

One selection technique involves incorporating into the expression vector a DNA sequence marker, with any necessary control elements, that codes for a selectable trait in the transformed cell. These markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture, and tetracyclin, kanamycin or ampicillin (i.e. β-lactamase) resistance genes for culturing in *E.coli* and other bacteria. Alternatively, the gene for such selectable trait can be on another vector, which is used to co-transform the desired host cell.

Another method of identifying successfully transformed cells involves growing the cells resulting from the introduction of a plasmid of the invention, optionally to allow the expression of a recombinant polypeptide (i.e. a polypeptide which is encoded by a polynucleotide sequence on the plasmid and is heterologous to the host cell, in the sense that that polypeptide is not naturally produced by the host). Cells can be harvested and lysed and their DNA or RNA content examined for the presence of the recombinant sequence using a method such as that described by Southern (1975) *J. Mol. Biol.* **98**, 503 or Berent *et al* (1985) *Biotech.* **3**, 208 or other methods of DNA and RNA analysis common in the art. Alternatively, the presence of a polypeptide in the supernatant of a culture of a transformed cell can be detected using antibodies.

In addition to directly assaying for the presence of recombinant DNA, successful transformation can be confirmed by well known immunological methods when the recombinant DNA is capable of directing the expression of the protein. For example, cells successfully transformed with an expression vector produce proteins displaying appropriate antigenicity. Samples of cells

suspected of being transformed are harvested and assayed for the protein using suitable antibodies.

Thus, in addition to the transformed host cells themselves, the present invention also contemplates a culture of those cells, preferably a monoclonal (clonally homogeneous) culture, or a culture derived from a monoclonal culture, in a nutrient medium.

Transformed host cells may be cultured for a sufficient time and under appropriate conditions known to those skilled in the art, and in view of the teachings disclosed herein, to permit the expression of the chaperone and heterologous protein encoded by the plasmid.

The culture medium may be non-selective or place a selective pressure on the maintenance of the plasmid.

The thus produced heterologous protein may be present intracellularly or, if secreted, in the culture medium and/or periplasmic space of the host cell.

The step of "purifying the thus expressed heterologous protein from the cultured host cell or the culture medium" optionally comprises cell immobilization, cell separation and/or cell breakage, but always comprises at least one other purification step different from the step or steps of cell immobilization, separation and/or breakage.

Cell immobilization techniques, such as encasing the cells using calcium alginate bead, are well known in the art. Similarly, cell separation techniques, such as centrifugation, filtration (e.g. cross-flow filtration, expanded bed chromatography and the like are well known in the art.

Likewise, methods of cell breakage, including beadmilling, sonication, enzymatic exposure and the like are well known in the art.

The at least one other purification step may be any other step suitable for protein purification known in the art. For example purification techniques for the recovery of recombinantly expressed albumin have been disclosed in: WO 92/04367, removal of matrix-derived dye; EP 464 590, removal of yeast-derived colorants; EP 319 067, alkaline precipitation and subsequent application of the albumin to a lipophilic phase; and WO 96/37515, US 5 728 553 and WO 00/44772, which describe complete purification processes; all of which are incorporated herein by reference.

Proteins other than albumin may be purified from the culture medium by any technique that has been found to be useful for purifying such proteins.

Suitable methods include ammonium sulphate or ethanol precipitation, acid or solvent extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, concentration, dilution, pH adjustment, diafiltration, ultrafiltration, high performance liquid chromatography ("HPLC"), reverse phase HPLC, conductivity adjustment and the like.

Thus, the heterologous protein can be separated from contaminating molecules and/or concentrated. It is preferred that the purified heterologous protein is provided at a concentration of at least 0.03 $g.L^{-1}$, 0.04 $g.L^{-1}$, 0.05 $g.L^{-1}$,0.06 $g.L^{-1}$,0.07 $g.L^{-1}$, 0.08 $g.L^{-1}$, 0.09 $g.L^{-1}$, 0.1 $g.L^{-1}$, 0.2 $g.L^{-1}$, 0.3 $g.L^{-1}$, 0.4 $g.L^{-1}$, 0.5 $g.L^{-1}$, 0.6 $g.L^{-1}$, 0.7 $g.L^{-1}$, 0.8 $g.L^{-1}$, 0.9 $g.L^{-1}$, 1 $g.L^{-1}$, 2 $g.L^{-1}$, 3 $g.L^{-1}$, 4 $g.L^{-1}$, 5 $g.L^{-1}$, 6 $g.L^{-1}$, 7 $g.L^{-1}$, 8 $g.L^{-1}$, 9 $g.L^{-1}$, 10 $g.L^{-1}$, 15 $g.L^{-1}$, 20 g.L

^{1}, 25 $g.L^{-1}$, 30 $g.L^{-1}$, 40 $g.L^{-1}$,50 $g.L^{-1}$, 60 $g.L^{-1}$, 70 $g.L^{-1}$, 70 $g.L^{-1}$, 90 $g.L^{-1}$, 100 $g.L^{-1}$, 150 $g.L^{-1}$, 200 $g.L^{-1}$,250 $g.L^{-1}$, 300 $g.L^{-1}$, 350 $g.L^{-1}$, 400 $g.L^{-1}$, 500 $g.L^{-1}$, 600 $g.L^{-1}$, 700 $g.L^{-1}$, 800 $g.L^{-1}$, 900 $g.L^{-1}$, 1000 $g.L^{-1}$, or more.

It is preferred that the heterologous protein is purified to achieve a pharmaceutically acceptable level of purity. A protein has a pharmaceutically acceptable level of purity is it is essentially pyrogen free and can be administered in a pharmaceutically efficacious amount without causing medical effects not associated with the activity of the protein.

The resulting heterologous protein may be used for any of its known utilities, which, in the case of albumin, include i.v. administration to patients to treat severe burns, shock and blood loss, supplementing culture media, and as an excipient in formulations of other proteins.

Although it is possible for a therapeutically useful heterologous protein obtained by a process of the of the invention to be administered alone, it is preferable to present it as a pharmaceutical formulation, together with one or more acceptable carriers or diluents. The carrier(s) or diluent(s) must be "acceptable" in the sense of being compatible with the desired protein and not deleterious to the recipients thereof. Typically, the carriers or diluents will be water or saline which will be sterile and pyrogen free.

Optionally the thus formulated protein will be presented in a unit dosage form, such as in the form of a tablet, capsule, injectable solution or the like.

A further embodiment of the present invention provides a host cell recombinantly encoding proteins comprising the sequences of PDI and transferrin-based proteins. By "transferrin-based protein" we mean

transferrin or any other member of the transferrin family (e.g. lactoferrin), a variant or fragment thereof or a fusion protein comprising transferrin, a variant or fragment thereof, including the types described above. Thus the present invention also provides for the use of a recombinant PDI gene to increase the expression of a transferrin-based protein.

The PDI gene may be provided on a plasmid, such as a 2µm-family plasmid as described above. Alternatively, the PDI gene may be chromosomally integrated. In a preferred embodiment, the PDI gene is chromosomally integrated at the locus of an endogenously encoded PDI gene, preferably without disrupting the expression of the endogenous PDI gene.

The gene encoding the transferrin-based protein may be provided on a plasmid, such as a 2µm-family plasmid as described above, or may be chromosomally integrated, such as at the locus of an endogenously encoded PDI gene, preferably without disrupting the expression of the endogenous PDI gene.

In one embodiment the PDI gene is chromosomally integrated and the gene encoding the transferrin-based protein is provided on a plasmid. In another embodiment, the PDI gene is provided on a plasmid and the gene encoding the transferrin-based protein is chromosomally integrated. In another embodiment both the PDI gene and the gene encoding the transferrin-based protein are chromosomally integrated. In another embodiment both the PDI gene and the gene encoding the transferrin-based protein are provided on a plasmid.

As is apparent from the examples of the present application, the combination of recombinantly expressed PDI and transferrin-based proteins

provides a surprisingly high level of transferrin expression. For example, transferrin expression in a system that includes a chromosomally encoded recombinant PDI gene provided a 2-fold increase (compared to a control in which there is no chromosomally encoded recombinant PDI gene). This increase was 5-times greater than an equivalent system comprising a recombinant gene encoding human albumin in place of the recombinant transferrin gene.

The host may be any cell type, such as a prokaryotic cell (e.g. bacterial cells such as *E. coli*) or a eukaryotic cell. Preferred eukaryotic cells include fungal cells, such as yeast cells, and mammalian cells. Exemplary yeast cells are discussed above. Exemplary mammalian cells include human cells.

Host cells as described above can be cultured to produce recombinant transferrin-based proteins. The thus produced transferrin-based proteins can be isolated from the culture and purified, preferably to a pharmaceutically acceptable level of purity, for example using techniques known in the art and/or as set out above. Purified transferrin-based proteins may be formulated with a pharmaceutically acceptable carrier or diluent and may be presented in unit dosage form.

The present invention will now be exemplified with reference to the following non-limiting examples and figures.

BRIEF DESCRIPTION OF THE FIGURES

Figures 1, 2, 4, 6 to 15, 22 and 25 show various plasmid maps.

Figure 3 shows plasmid insertion sites.

Figure 5 shows a restriction map of a DNA fragment containing the PDI coding sequence.

Figure 16 shows the results of rocket immunoelectrophoresis (RIE) determination of increased recombinant transferrin (N413Q, N611Q) secretion with *PDI1* over-expression. Cryopreserved yeast stocks were grown for 4-days in 10mL BMMD shake flask cultures and supernatants were loaded at 5μL per well. Goat polyclonal anti-transferrin (human) antiserum (Calbiochem) was used at 40μL per rocket immunoelectrophoresis gel (50mL). A = Control strain [pSAC35], duplicate flasks; B = Control strain [pDB2536], duplicate flasks; C = Control strain [pDB2711], neat to 40-fold aqueous dilutions; D = Control strain [pDB2931], duplicate flasks; E = Control strain [pDB2929], neat to 40-fold aqueous dilutions.

Figure 17 shows the results of RIE analysis of recombinant transferrin (N413Q, N611Q) secretion with and without *PDI1* over-expression. Cryopreserved yeast stocks were grown for 4-days in 10mL BMMD shake flask cultures and supernatants were loaded at 5μL per well. Duplicate loadings were made of supernatants from two individual cultures of each strain. Goat polyclonal anti-transferrin (human) antiserum (Calbiochem) was used at 40μL per rocket immunoelectrophoresis gel (50mL). A =

Control strain [pSAC35]; B = Control strain [pDB2536]; C = Control strain [pDB2711]; D = Control strain [pDB2931]; E = Control strain [pDB2929].

Figure 18 shows the results of SDS-PAGE analysis of recombinant transferrin secretion with and without *PDI1* over-expression. BMMD shake flask cultures were grown for 4-days and 10μL supernatant analysed on non-reducing SDS-PAGE (4-12% NuPAGE®, MOPS buffer, InVitrogen) with GelCode® Blue reagent (Pierce). SeeBlue Plus2 Markers (InVitrogen). 1 = pDB2536; 2 = pDB2536; 3 = pDB2711; 4 = pDB2711; 5 = pDB2931; 6 = pDB2931; 7 = pDB2929; 8 = pDB2929; 9 = pSAC35 control.

Figure 19 shows RIE analysis of recombinant transferrin secretion from *S. cerevisiae* strains with an additional integrated copy of *PDI1*. 5-day BMMD shake flask culture supernatants were loaded at 5mL per well. Strains contained; 1) pSAC35 (negative control); 2) pDB2536 (recombinant non-glycosylated transferrin (N413Q, N611Q)) or 3) pDB2506 (same as plasmid pDB2536 but the transferrin ORF encodes transferrin without the N→Q mutations at positions 413 and 611, i.e. recombinant glycosylated transferrin). Each well contained a sample derived from an individual transformant. Standards were human plasma holo-transferrin (Calbiochem) at 100, 50, 20, 10, 5 and $2mg.L^{-1}$.

Figure 20 shows RIE analysis of recombinant transferrin secretion from Strain A [pDB2536] and Strain A [pDB2506] grown in shake flask culture. 5-day BMMD shake flask culture supernatants were loaded in duplicate at 5mL per well.

Figure 21 shows SDS-PAGE analysis of recombinant transferrin secreted from Strain A [pDB2536] and Strain A [pDB2506] grown in shake flask

culture. Cultures were grown for 5-days in BMMD and 30mL supernatants analysed on SDS-PAGE (4-12% NuPAGE™, MOPS Buffer, InVitrogen) stained with GelCode, Blue Reagent (Pierce). 1) Strain A [pDB2536] transformant 1; 2) Strain A [pDB2536] transformant 2; 3) Strain A [pSAC35] control; 4) Strain A [pDB2506] transformant 1; 5) SeeBlue, Plus2 Protein Standards (approximate molecular weights only).

Figure 23 shows RIE of recombinant transferrin secreted from *S. cerevisiae* strains with different PDI1 copy numbers. 3-day BMMD shake flask culture supernatants were loaded at 5mL per well. Goat polyclonal anti-transferrin (human) antiserum (Calbiochem) was used at 30mL per rocket immunoelectrophoresis gel (50mL). (A) supernatant from *S. cerevisiae* control strain [pDB2711] or [pDB2712]; (B) supernatant from Strain A [pDB2536]; (C) supernatant from control strain [pDB2536].

Figure 24 shows SDS-PAGE analysis of recombinant transferrin secreted from *S. cerevisiae* strains with different PDI1 copy numbers. 4-12% NuPAGE reducing gel run with MOPS buffer (InVitrogen) after loading with 30mL of 3-day BMMD shake flask culture supernatant per lane; (lane 1) supernatant from control strain [pDB2536]; (lane 2) supernatant from Strain A [pDB2536]; (lanes 3-6) supernatant from control strain [pDB2711] or [pDB2712]; (lane 7) molecular weight markers (SeeBlue Plus2, InVitrogen).

EXAMPLES

Two types of expression cassette have been used to secrete a recombinant human transferrin mutant (N413Q, N611Q) from *S. cerevisiae*. One type uses a modified HSA(pre)/MFα1(pro) leader sequence (named the

"modified fusion leader" sequence). The second type of expression cassette uses only the modified HSA(pre) leader sequence.

The 24 amino acid sequence of the "modified fusion leader" is MKWVFIVSILFLFSSAYSRSLDKR.

The 18 amino acid sequence of the modified HSA(pre) leader sequence is MKWVFIVSILFLFSSAYS.

Transferrin (N413Q, N611Q) expression using these two cassettes has been studied in *S. cerevisiae* using the 2μm expression vector with and without an additional copy of the *S. cerevisiae* PDI gene, *PDI1*.

<u>Example 1: Construction of expression plasmids</u>

A 52-bp linker made by annealing 0.5mM solutions of oligonucleotides CF86 and CF87 (see below) was introduced into the US-region of the 2μm plasmid pSAC35 at the *Xcm*I-sites in the 599-bp inverted repeats. One *Xcm*I-site cuts 51-bp after the *REP2* translation termination codon, whereas the other *Xcm*I-site cuts 127-bp before the end of the *FLP* coding sequence, due to overlap with the inverted repeat (see Figure 3). This DNA linker contained a core region "*Sna*BI-*Pac*I-*Fse*I/*Sfi*I-*Sma*I-*Sna*BI", which encoded restriction sites absent from pSAC35.

<u>*Xcm*I Linker (CF86+CF87)</u>

```
                                    SfiI
                               --------------
                       PacI                          SnaBI
                    --  .  ---                      -------
              SnaBI               FseI       SmaI
             -------           --------      -- . --
CF86  GGAGTGGTA CGTATTAATT AAGGCCGGCC AGGCCCGGGT ACGTACCAAT TGA
CF87 TCCTCACCAT GCATAATTAA TTCCGGCCGG TCCGGGCCCA TGCATGGTTA AC
```

Plasmid pSAC35 was partially digested with *Xcm*I, the linear 11-kb fragment was isolated from a 0.7%(w/v) agarose gel, ligated with the CF86/CF87 *Xcm*I linker (neat, 10^{-1} and 10^{-2} dilutions) and transformed into *E. coli* DH5α. Ampicillin resistant transformants were selected and screened for the presence of plasmids that could be linearised by *Sma*I digestion. Restriction enzyme analysis identified pDB2688 (Figure 4) with the linker cloned into the *Xcm*I-site after *REP2*. DNA sequencing using oligonucleotides primers CF88, CF98 and CF99 (Table 1) confirmed the insertion contained the correct linker sequence.

Table 1

Oligonucleotide sequencing primers:

Primer	Description	Sequence
CF88	*REP2* primer, 20mer	5'-ATCACGTAATACTTCTAGGG-3'
CF98	*REP2* primer, 20mer	5'-AGAGTGAGTTGGAAGGAAGG-3'
CF99	*REP2* primer, 20mer	5'-AGCTCGTAAGCGTCGTTACC-3'

The yeast strain was transformed to leucine prototrophy using a modified lithium acetate method (Sigma yeast transformation kit, YEAST-1, protocol 2; (Ito *et al*, 1983, *J. Bacteriol.*, 153, 163; Elble, 1992, *Biotechniques*, 13, 18)). Transformants were selected on BMMD-agar plates, and were subsequently patched out on BMMD-agar plates. Cryopreserved trehalose stocks were prepared from 10mL BMMD shake flask cultures (24 hrs, 30^0C, 200rpm).

The composition of YEPD and BMMD is described by Sleep *et al.*, 2002, *Yeast*, **18**, 403. YEPS and BMMS are similar in composition to YEPD and BMMD accept that 2% (w/v) sucrose was substituted for the 2% (w/v) glucose as the sole initial carbon source.

The *S. cerevisiae PDI1* gene was cloned into the *Xcm*I-linker of pDB2688. The *PDI1* gene (Figure 5) was cloned on a 1.9-kb *Sac*I-*Spe*I fragment from a larger *S. cerevisiae* genomic DNA fragment containing the *PDI1* gene, which had been cloned into YIplac211 (Gietz & Sugino, 1988, *Gene*, **74**, 527-534), which had been treated with T4 DNA polymerase to fill the *Spe*I 5'-overhang and remove the *Sac*I 3'-overhang. This *PDI1* fragment included 212-bp of the *PDI1* promoter upstream of the translation initiation codon, and 148-bp downstream of the translation termination codon. This was ligated with *Sma*I linearised/calf intestinal alkaline phosphatase treated pDB2688, to create plasmid pDB2690 (Figure 6), with the *PDI1* gene transcribed in the same direction as *REP2*. A *S. cerevisiae* strain was transformed to leucine prototrophy with pDB2690.

An expression cassette for a human transferrin mutant (N413Q, N611Q) was subsequently cloned into the *Not*I-site of pDB2690 to create pDB2711 (Figure 7). The expression cassette in pDB2711 contains the *S. cerevisiae PRB1* promoter, an HSA/MFα fusion leader sequence (EP 387319; Sleep *et al*, 1990, *Biotechnology (N.Y.)*, **8**, 42) followed by a coding sequence for the human transferrin mutant (N413Q, N611Q) and the *S. cerevisiae ADH1* terminator. Plasmid pDB2536 was constructed similarly by insertion of the same expression cassette into the *Not*I-site of pSAC35.

The "modified fusion leader" sequence used in pDB2536 and pDB2711 comprises a modified HSA-pre sequence and a MFα1-pro sequence. An alternative leader sequence used was the modified HSA-pre sequence, which was derived from the modified fusion leader sequence by removal of the six residues of the MFα1-pro sequence.

The modified fusion leader sequence in pDB2515 (Figure 8) was mutated with oligonucleotides CF154 and CF155 to delete the coding sequence for the six residues (RSLDKR) of the MFα1-pro region. This was performed according to the instruction manual of the Statagene's QuickChange™ Site-Directed Mutagenesis Kit. pDB2515 is the *E. coli* cloning vector pGEM-7Z(-) (Promega) containing the 2940-bp *Not*I-*Hin*dIII (partial) DNA fragment of pDB2529 (see below) ligated between the *Psp*OMI and *Hin*dIII sites.

CF154

5'-GTTCTTGTTCTCCTCTGCTTACTCTGTCCCTGATAAAACTGTGAGA TGG-3'

CF155

5'-CCATCTCACAGTTTTATCAGGGACAGAGTAAGCAGAGGAGAACAA GAAC-3'

Competent *E. coli* DH5α cells were transformed with the mutated plasmids and ampicillin resistant colonies were selected. Plasmid DNA from these colonies was screened by double digestion with *Eco*RI and *Bgl*II. The correct DNA sequence for the modified HSA-pre leader was subsequently confirmed in pDB2921 (Figure 9) over a 386-bp region between the *Afl*II and *Bam*HI sites either side of the leader sequence. This 386-bp *Afl*II-

*Bam*HI fragment was isolated, and ligated with a 6,081-bp *Afl*II-*Bam*HI fragment from pDB2529 (Figure 10), prepared by partial digestion with *Bam*HI and complete digestion with *Afl*II and calf intestinal alkaline phosphatase. pDB2529 is the *E. coli* cloning vector pBST(+) (Sleep *et al*, 2001, *Yeast*, 18, 403-441) containing the transferrin expression cassette of pDB2536 cloned into the unique *Not*I-site. This produced pDB2928 (Figure 11), which was isolated from ampicillin resistant *E. coli* DH5α cells transformed with the ligation products.

The 3,256-bp *Not*I expression cassette was isolated from pDB2928. This contained the *PRB1* promoter, the coding region for the modified HSA-pre leader sequence followed by transferrin (N413Q, N611Q), and the *ADH1* terminator. This was ligated into the *Not*I sites of the 2μm-based vectors pSAC35 and pDB2690 to generate the expression plasmids pDB2929, pDB2930, pDB2931 and pDB2932 (Figures 12-15). In pDB2929 and pDB2931 the transferrin (N413Q, N611Q) sequence is transcribed in the same direction as *LEU2*, whereas in pDB2930 and pDB2932 transcription is in the opposite direction.

Example 2: Expression of transferrin

A *S. cerevisiae* control strain was transformed to leucine prototrophy with all the transferrin (N413Q, N611Q) expression cassettes, and cryopreserved stocks were prepared.

Strains were grown for four days at 30°C in 10mL BMMD cultures in 50mL conical flasks shaken at 200rpm. The titres of recombinant transferrin secreted into the culture supernatants were compared by rocket immunoelectrophoresis (RIE), reverse phase high performance liquid

chromatography (RP-HPLC) (Table 2), and non-reducing SDS polyacrylamide electrophoresis stained with colloidal Coomassie blue stain (SDS-PAGE). The increase in recombinant transferrin secreted when *S. cerevisiae PDI1* was over-expressed was estimated to be greater than 10-fold.

Table 2:

Plasmid	Secretory Leader	Additional *PDI1*	Average Transferrin Titre ($\mu g.mL^{-1}$) (n=2)	Estimated Increase due to Additional *PDI1*
pSAC35	None	No	0.4	-
pDB2536	Fusion Leader	No	6.2	-
pDB2711	Fusion Leader	Yes	112.8	18-fold
pDB2931	Modified HSA-pre Leader	No	5.1	-
pDB2929	Modified HSA-pre Leader	Yes	76.1	15-fold

RIE analysis indicated that the increased transferrin secretion in the presence of additional copies of *PDI1* was approximately 15-fold (Figure 16). By RIE analysis the increase appeared slightly larger for the modified

HSA-pre leader sequence than for the modified fusion leader sequence (Figure 17).

By RP-HPLC analysis the increase in transferrin secretion was determined to be 18-fold for the modified fusion leader sequence and 15-fold for the modified HSA-pre leader sequence (Table 2).

Figure 18 shows an SDS-PAGE comparison of the recombinant transferrin secreted by *S. cerevisiae* strains with and without additional *PDI1* expression.

RP-HPLC Method for Determining Transferrin Expression

Column: 50 × 4.6mm Phenomenex Jupiter C4 300Å, 5μm

Column temperature: 45°C

Flow rate: $1 mL.min^{-1}$

Peak detection:UV absorbance at 214nm

HPLC mobile phase A: 0.1% TFA, 5% Acetonitrile

HPLC mobile phase B: 0.1% TFA, 95% Acetonitrile

Gradient: 0 to 3 minutes 30% B
3 to 13 minutes 30 to 55% B in a linear gradient
13 to 14 minutes 55% B
14 to 15 minutes 55 to 30% B in a linear gradient
15 to 20 minutes 30% B

Injection: Generally 100μL of sample, but any volume can be injected

Standard Curve: 0.1 to 10μg of human transferrin injected vs peak area

Standard curve used for the results shown was linear up to 10μg.

$$y = 530888.x + 10526.7$$

where y = peak area, and x = amount in μg.

(r^2): 0.999953, where Correlation Coefficient = r

Example 3:

S. cerevisiae Strain A was selected to investigate the secretion of recombinant glycosylated transferrin expression from plasmid pDB2506 and recombinant non-glycosylated transferrin (N413Q, N611Q) from plasmid pDB2536. Strain A has the following characteristics –

- additional chromosomally integrated *PDI1* gene integrated at the host *PDI1* chromosomal location.

- the *URA3* gene and bacterial DNA sequences containing the ampicillin resistance gene were also integrated into the *S. cerevisiae* genome at the insertion sites for the above genes.

A control strain had none of the above insertions.

Control strain [cir^0] and Strain A [cir^0] were transformed to leucine prototrophy with pDB2506 (recombinant transferrin), pDB2536 (recombinant non-glycosylated transferrin (N413Q, N611Q)) or pSAC35 (control). Transformants were selected on BMMD-agar.

The relative level of transferrin secretion in BMMD shake flask culture was determined for each strain/plasmid combination by rocket immunoelectrophoresis (RIE). Figure 19 shows that both strains secreted both the glycosylated and non-glycosylated recombinant transferrins into the culture supernatant.

The levels of both the glycosylated and non-glycosylated transferrins secreted from Strain A [pDB2506] and Strain A [pDB2536] respectively, appeared higher than the levels secreted from the control strain. Hence, at least in shake flask culture, *PDI1* integrated into the host genome at the *PDI1* locus in Strain A has enhanced transferrin secretion.

Furthermore, the increase in transferrin secretion observed between control strain [pDB2536] and Strain A [pDB2536] appeared to be at least a 100% increase by RIE. In contrast, the increase in rHA monomer secretion between control strain [pDB2305] and Strain A [pDB2305] was approximately 20% (data not shown). Therefore, the increase in transferrin secretion due to the additional copy of *PDI1* in Strain A was surprising large considering that transferrin has 19 disulphide bonds, compared to rHA with 17 disulphide bonds. Additional copies of the *PDI1* gene may be particularly beneficial for the secretion from *S. cerevisiae* of proteins from the transferrin family, and their derivatives.

The levels of transferrin secreted from Strain A [pDB2536] and Strain A [pDB2506] were compared by RIE for transformants grown in BMMD and YEPD (Figure 20). Results indicated that a greater than 2-fold increase in titres of both non-glycosylated recombinant transferrin (N413Q, N611Q) and glycosylated recombinant transferrin was achieved by growth in YEPD (10-20 $mg.L^{-1}$ serum transferrin equivalent) compared to BMMD (2-5 $mg.L^{-1}$ serum transferrin equivalent). The increase in both glycosylated and non-glycosylated transferrin titre observed in YEPD suggested that both expression transferrin plasmids were sufficiently stable under non-selective growth conditions to allow the expected increased biomass which usually results from growth in YEPD to be translated into increased glycosylated and non-glycosylated transferrin productivity.

SDS-PAGE analysis of non-glycosylated transferrin (N413Q, N611Q) secreted from Strain A [pDB2536] and glycosylated transferrin from Strain A [pDB2506] grown in BMMD shake flask culture is shown in Figure 21. Strain A [pDB2536] samples clearly showed an additional protein band compared to the Strain A [pSAC35] control. This extra band migrated at the expected position for the recombinant transferrin (N413Q, N611Q) secreted from control strain [pDB2536]. Strain A [pDB2506] culture supernatants appeared to contain a diffuse protein band at the position expected for transferrin. This suggested that the secreted recombinant transferrin was heterogeneous, possibly due to hyper-manosylation at Asp413 and/or Asp611.

Example 4: Comparing transferrin secretion from *S. cerevisiae* control strain containing pDB2711 with transferrin secretion from *S. cerevisiae* Strain A

Plasmid pDB27111 is as described above. Plasmid pDB2712 (Figure 22) was also produced with the *Not*I cassette in the opposite direction to pDB2711.

Control strain *S. cerevisiae* [cir^0] was transformed to leucine prototrophy with pDB2711 and pDB2712. Transformants were selected on BMMD-agar and cryopreserved trehalose stocks of control strain [pDB2711] were prepared (Table 3).

Table 3: Recombinant transferrin titres from high cell density fermentations

Strain	Supernatant (g.L^{-1})	
	GP-HPLC	SDS-PAGE
Control [pDB2536]	0.5/0.4	-
Alternative control [pDB2536]	1.5/1.6	0.6
	0.9/0.9	0.4/0.4/0.5
Strain A [pDB2536]	0.7	0.6
	0.6	-
Control [pDB2711]	3.5	3.6
	3.4	2.7/3.1

Secretion of recombinant transferrin (N413Q, N611Q) by control strain [pDB2711], control strain [pDB2712], Strain A [pDB2536], control strain [pDB2536] and an alternative control strain [pDB2536] was compared in both BMMD and YEPD shake flask culture. RIE indicated that a significant increase in recombinant transferrin secretion had been achieved from control strain [pDB2711] with multiple episomal *PDI1* copies, compared to Strain A [pDB2536] with two chromosomal copies of *PDI1*, and control strain [pDB2536] with a single chromosomal copy of *PDI1* gene (Figure 23). Control strain [pDB2711] and control strain [pDB2712] appeared to secrete similar levels of rTf (N413Q, N611Q) into the culture media. The levels of secretion were relatively consistent between control strain [pDB2711] and control strain [pDB2712] transformants in both BMMD and YEPD media, suggesting that plasmid stability was sufficient

for high-level transferrin secretion even under non-selective conditions. This is in contrast to the previous published data in relation to recombinant PDGF-BB and HSA where introduction of PDI1 into multicopy 2μm plasmids was shown to be detrimental to the host.

Reducing SDS-PAGE analysis of transferrin secreted from control strain [pDB2711], control strain [pDB2712], Strain A [pDB2536], control strain [pDB2536] and alternative control strain [pDB2536] in BMMD shake flask culture is shown in Figure 24. This shows an abundant protein band in all samples from control strain [pDB2711] and control strain [pDB2712] at the position expected for transferrin (N413Q, N611Q). The relative stain intensity of the transferrin (N413Q, N611Q) band from the different strains suggested that Strain A [pDB2536] produced more than control strain [pDB2536] and alternative control strain [pDB2536], but that there was an even more dramatic increase in secretion from control strain [pDB2711] and control strain [pDB2712]. The increased recombinant transferrin secretion observed was concomitant with the increased *PDI1* copy number in these strains. This suggested that Pdi1p levels were limiting transferrin secretion in control strain, Strain A and the alternative control strain, and that elevated *PDI1* copy number was responsible for increased transferrin secretion. Elevated *PDI1* copy number could increase the steady state expression level of *PDI1* so increasing the amount of Pdi1p activity. There are a number of alternative methods by which this could be achieved without increasing the copy number of the *PDI1* gene, for example the steady state *PDI1* mRNA level could be increased by either increasing the transcription rate, say by use of a higher efficiency promoter, or by reducing the clearance rate of the *PDI1* mRNA. Alternatively, protein engineering could be used to enhance the specific activity or turnover number of the Pdi1p protein.

In high cell density fermentations control strain [pDB2711] recombinant transferrin (N413Q, N611Q) production was measured at approximately 3g.L^{-1} by both GP-HPLC analysis and SDS-PAGE analysis (Table 3). This level of production is several fold-higher than control strain, the alternative control strain or Strain A containing pDB2536. Furthermore, for the production of proteins for therapeutic use in humans, expression systems such as control strain [pDB2711] have advantages over those using Strain A, as they do not contain bacterial DNA sequences.

CONCLUSIONS

Secretion of recombinant transferrin from a multicopy expression plasmid (pDB2536) is investigated in *S. cerevisiae* strains containing an additional copy of the *PDI1* gene integrated into the yeast genome. Transferrin secretion was also investigated from in *S. cerevisiae* transformed with a multicopy expression plasmid, in which the *PDI1* gene has been inserted into the multicopy episomal transferrin expression plasmid in which *PDI1* is also inserted (pDB2711).

A *S. cerevisiae* strain with an additional copy of the *PDI1* gene integrated into the genome at the endogenous *PDI1* locus, secreted recombinant transferrin and non-glycosylated recombinant transferrin (N413Q, N611Q) at an elevated level compared to strains containing a single copy of *PDI1*. A further increase in *PDI1* copy number was achieved by using pDB2711 In high cell density fermentation of the strain transformed with pDB2711, recombinant transferrin (N413Q, N611Q) was secreted at approximately 3g.L^{-1}, as measured by SDS-PAGE and GP-HPLC analysis. Therefore,

increased *PDI1* gene copy number has produced a large increase in the quantity of recombinant transferrins secreted from *S. cerevisiae*.

The following conclusions are drawn –

1. In shake flask analysis of recombinant transferrin expression from pDB2536 (non-glycosylated transferrin (N413Q, N611Q) and pDB2506 (glycosylated transferrin) the *S. cerevisiae* strain Strain A secreted higher levels of both recombinant transferrins into the culture supernatant than control strains. This was attributed to the extra copy of *PDI1* integrated at the *PDI1* locus.

2. Control strain [pDB2711], which contained the *PDI1* gene on the multicopy expression plasmid, produced a several-fold increase in recombinant transferrin (N413Q, N611Q) secretion compared to Strain A [pDB2536] in both shake flask culture and high cell density fermentation.

3. Elevated *PDI1* copy number in yeast such as *S. cerevisiae* will be advantageous during the production of heterologous proteins, such as those from the transferrin family.

4. pSAC35-based plasmids containing additional copies of *PDI1* gene have advantages for the production of proteins from the transferrin family, and their derivatives, such as fusions, mutants, domains and truncated forms.

CLAIMS

1. A method for producing heterologous protein comprising:

(a) providing a host cell comprising a 2μm-family plasmid, the plasmid comprising a gene encoding protein comprising the sequence of a chaperone protein and a gene encoding a heterologous protein;

(b) culturing the host cell in a culture medium under conditions that allow the expression of the gene encoding protein comprising the sequence of the chaperone protein and the gene encoding a heterologous protein; and

(c) purifying the thus expressed heterologous protein from the cultured host cell or the culture medium; and

(d) optionally, lyophilising the thus purified protein.

2. The method of Claim 1 further comprising the step of formulating the purified heterologous protein with a carrier or diluent and optionally presenting the thus formulated protein in a unit dosage form.

3. Use of a 2μm-family plasmid as an expression vector to increase the production of a fungal (preferably yeast) or vertebrate heterologous protein by providing a gene encoding the heterologous protein and a gene encoding a chaperone protein on the same 2μm-family plasmid.

4. A 2μm-family plasmid comprising a gene encoding a protein comprising the sequence of a chaperone protein and a gene encoding a heterologous protein, wherein if the plasmid is based on the 2μm plasmid then it is a disintegration vector.

5. A method, use or plasmid according to any preceding claim wherein the chaperone has a sequence of a fungal chaperone (preferably a yeast chaperone) or a mammalian chaperone (preferably a human chaperone).

6. A method, use or plasmid according to any preceding claim wherein the chaperone is selected from *AHA1*, *CCT2*, *CCT3*, *CCT4*, *CCT5*, *CCT6*, *CCT7*, *CCT8*, *CNS1*, *CPR3*, *CPR6*, *EPS1*, *ERO1*, *EUG1*, *FMO1*, *HCH1*, *HSP10*, *HSP12*, *HSP104*, *HSP26*, *HSP30*, *HSP42*, *HSP60*, *HSP78*, *HSP82*, *JEM1*, *MDJ1*, *MDJ2*, *MPD1*, *MPD2*, *PDI1*, *PFD1*, *ABC1*, *APJ1*, *ATP11*, *ATP12*, *BTT1*, *CDC37*, *CPR7*, *HSC82*, *KAR2*, *LHS1*, *MGE1*, *MRS11*, *NOB1*, *ECM10*, *SSA1*, *SSA2*, *SSA3*, *SSA4*, *SSC1*, *SSE2*, *SIL1*, *SLS1*, *UBI4*, *ORM1*, *ORM2*, *PER1*, *PTC2*, *PSE1* and *HAC1* or truncated intronless *HAC1*.

7. A method, use or plasmid according to any preceding claim wherein the chaperone is protein disulphide isomerase.

8. A method, use or plasmid according to any preceding claim wherein the heterologous protein comprises a leader sequence effective to cause secretion in yeast.

9. A method, use or plasmid according to any preceding claim wherein the heterologous protein is a eukaryotic protein, or a fragment or

variant thereof, preferably a vertebrate or a fungal (preferably a yeast) protein.

10. A method, use or plasmid according to any preceding claim wherein the heterologous protein is a commercially useful protein.

11. A method, use or plasmid according to any preceding claim wherein the heterologous protein comprises a sequence selected from albumin, a monoclonal antibody, an etoposide, a serum protein (such as a blood clotting factor), antistasin, a tick anticoagulant peptide, transferrin, lactoferrin, endostatin, angiostatin, collagens, immunoglobulins, Fab' fragments, F(ab')2, scAb, scFv, interferons, interleukins, IL10, IL11, IL2, interferon α species and sub-species, interferon β species and sub-species, interferon γ species and sub-species, leptin, CNTF, $CNTF_{Ax15}$, IL1-receptor antagonist, erythropoietin (EPO) and EPO mimics, thrombopoietin (TPO) and TPO mimics, prosaptide, cyanovirin-N, 5-helix, T20 peptide, T1249 peptide, HIV gp41, HIV gp120, urokinase, prourokinase, tPA, hirudin, platelet derived growth factor, parathyroid hormone, proinsulin, insulin, glucagon, glucagon-like peptides, insulin-like growth factor, calcitonin, growth hormone, transforming growth factor β, tumour necrosis factor, G-CSF, GM-CSF, M-CSF, FGF, coagulation factors in both pre and active forms, including but not limited to plasminogen, fibrinogen, thrombin, pre-thrombin, pro-thrombin, von Willebrand's factor, α_1-antitrypsin, plasminogen activators, Factor VII, Factor VIII, Factor IX, Factor X and Factor XIII, nerve growth factor, LACI, platelet-derived endothelial cell growth factor (PD-ECGF), glucose oxidase, serum cholinesterase, aprotinin, amyloid precursor protein, inter-alpha trypsin inhibitor,

antithrombin III, apo-lipoprotein species, Protein C, Protein S, or a variant or fragment of any of the above.

12. A method, use or plasmid according to any preceding claim wherein the heterologous protein comprises the sequence of albumin or a variant or fragment thereof.

13. A method, use or plasmid according to any preceding claim wherein the heterologous protein comprises the sequence of a transferrin family member, preferably transferrin or lactoferrin, or a variant or fragment thereof.

14. A method, use or plasmid according to any preceding claim wherein the heterologous protein comprises a fusion protein, such as a fusion protein of albumin or a transferrin family member or a variant or fragment of either, fused directly or indirectly to the sequence of another protein.

15. A host cell comprising a plasmid as defined by any preceding claim.

16. A host cell according to Claim 15 which is a yeast cell.

17. A host cell according to Claim 16 in which the plasmid is based on pSR1, pSB3 or pSB4 and the yeast cell is *Zygosaccharomyces rouxii*, the plasmid is based on pSB1 or pSB2 and the yeast cell is *Zygosaccharomyces bailli*, , the plasmid is based on pSM1 and the yeast cell is *Zygosaccharomyces fermentati*, the plasmid is based on pKD1 and the yeast cell is *Kluyveromyces drosophilarum*, the plasmid is based on pPM1 and the yeast cell is *Pichia*

membranaefaciens, or the plasmid is based on the 2μm plasmid and the yeast cell is *Saccharomyces cerevisiae* or *Saccharomyces carlsbergensis*.

18. A host cell according to Claim 17 in which the plasmid is based on the 2μm plasmid and the yeast cell is *Saccharomyces cerevisiae* or *Saccharomyces carlsbergensis*.

The Patent Office

PATENTS • DESIGNS • COPYRIGHT • TRADE MARKS

CHARTER MARK

INVESTOR IN PEOPLE

Application No: GB 0329681.1
Claims searched: All

Examiner: Dr Rowena Dinham
Date of search: 29 March 2004

Patents Act 1977 : Search Report under Section 17

Documents considered to be relevant:

Category	Relevant to claims	Identity of document and passage or figure of particular relevance
A		*Plasmid*; Vol 25, pp 81-95 (1991). Ludwig & Bruschi. "*The 2μm plasmid as a nonselectable, stable, high copy number yeast plasmid*" See entire document, especially Results
A		*Gene*; Vol 132, pp 33-40 (1993). Ludwig *et al.* "*High-level heterologous gene expression in* Saccharomyces cerevisiae..." See entire document, especially Results and Discussion page 34 "*Vector and strain construction*"
A		US 5637504 A (HINCHLIFFE) See entire document, especially column 4 line 52-54, column 5 line 46-55 and Examples
A		WO 98/56928 A1 (STICHTING INSTITUT VOOR DIERHOUDERIJ EN DIERZONDHEID) See especially page 8 line 2-18 and Examples
A		US 5773245 A (WITTRUP) See especially column 3 line 32-60 and Examples

Categories:

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

Field of Search:

Search of GB, EP, WO & US patent documents classified in the following areas of the UKC^{W}:

Worldwide search of patent documents classified in the following areas of the IPC^{7}:

C12N

The following online and other databases have been used in the preparation of this search report:

WPI, EPODOC, JAPIO, MEDLINE, EMBASE, BIOSIS, SCISEARCH, CAPLUS

78

ABSTRACT

MODIFIED PLASMID AND USE THEREOF

The present invention provides a method for producing heterologous protein comprising:

(a) providing a host cell comprising a 2μm-family plasmid, the plasmid comprising a gene encoding a protein comprising the sequence of a chaperone protein and a gene encoding a heterologous protein;

(b) culturing the host cell in a culture medium under conditions that allow the expression of the gene encoding the chaperone protein and the gene encoding a heterologous protein; and

(c) purifying the thus expressed heterologous protein from the cultured host cell or the culture medium; and

(d) optionally, lyophilising the thus purified protein.

Figure 1

Figure 1

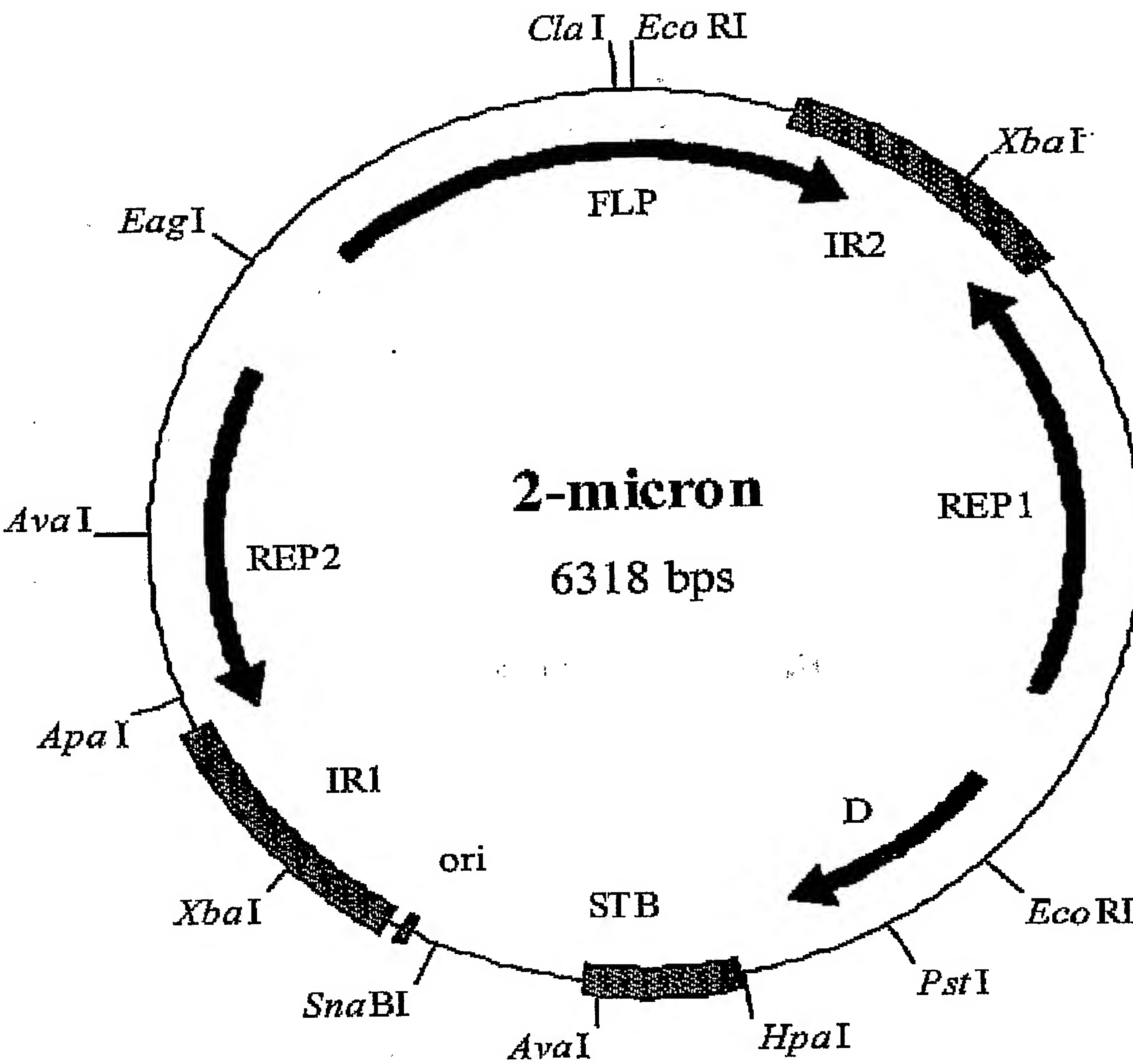

Cla I
Eco RI
Xba I
FLP
IR2
Eag I
REP1
2-micron
6318 bps
Ava I
REP2
Apa I
IR1
D
ori
STB
Xba I
Eco RI
Pst I
Sna BI
Ava I
Hpa I

Figure 2

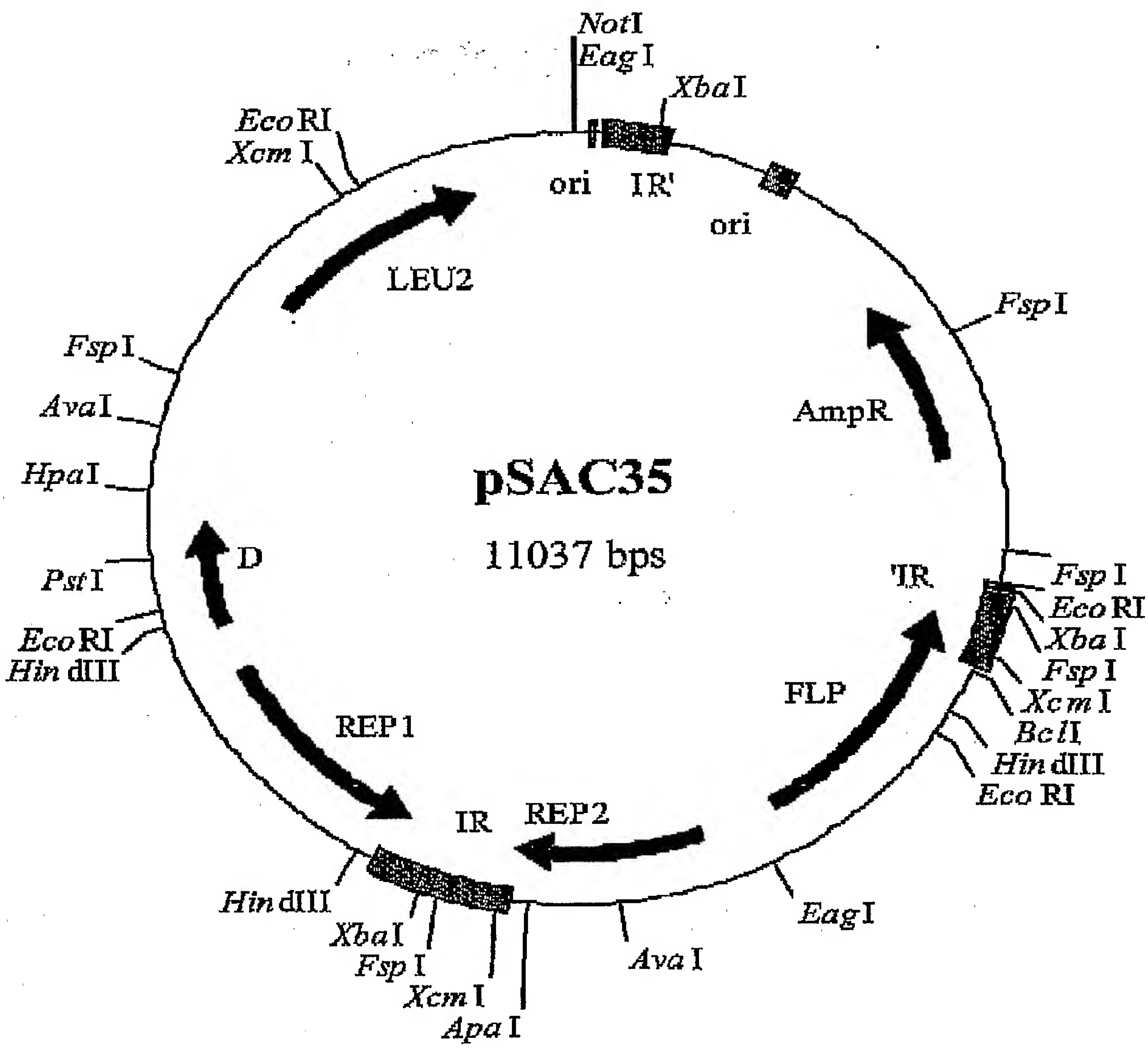

NotI
EagI
XbaI
EcoRI
XcmI
ori
IR'
ori
LEU2
FspI
AmpR
FspI
AvaI
HpaI
pSAC35
11037 bps
PstI
D
IR
FspI
EcoRI
XbaI
EcoRI
HindIII
FspI
FLP
XcmI
BclI
REP1
HindIII
EcoRI
IR
REP2
HindIII
EagI
XbaI
FspI
AvaI
XcmI
ApaI

Figure 3

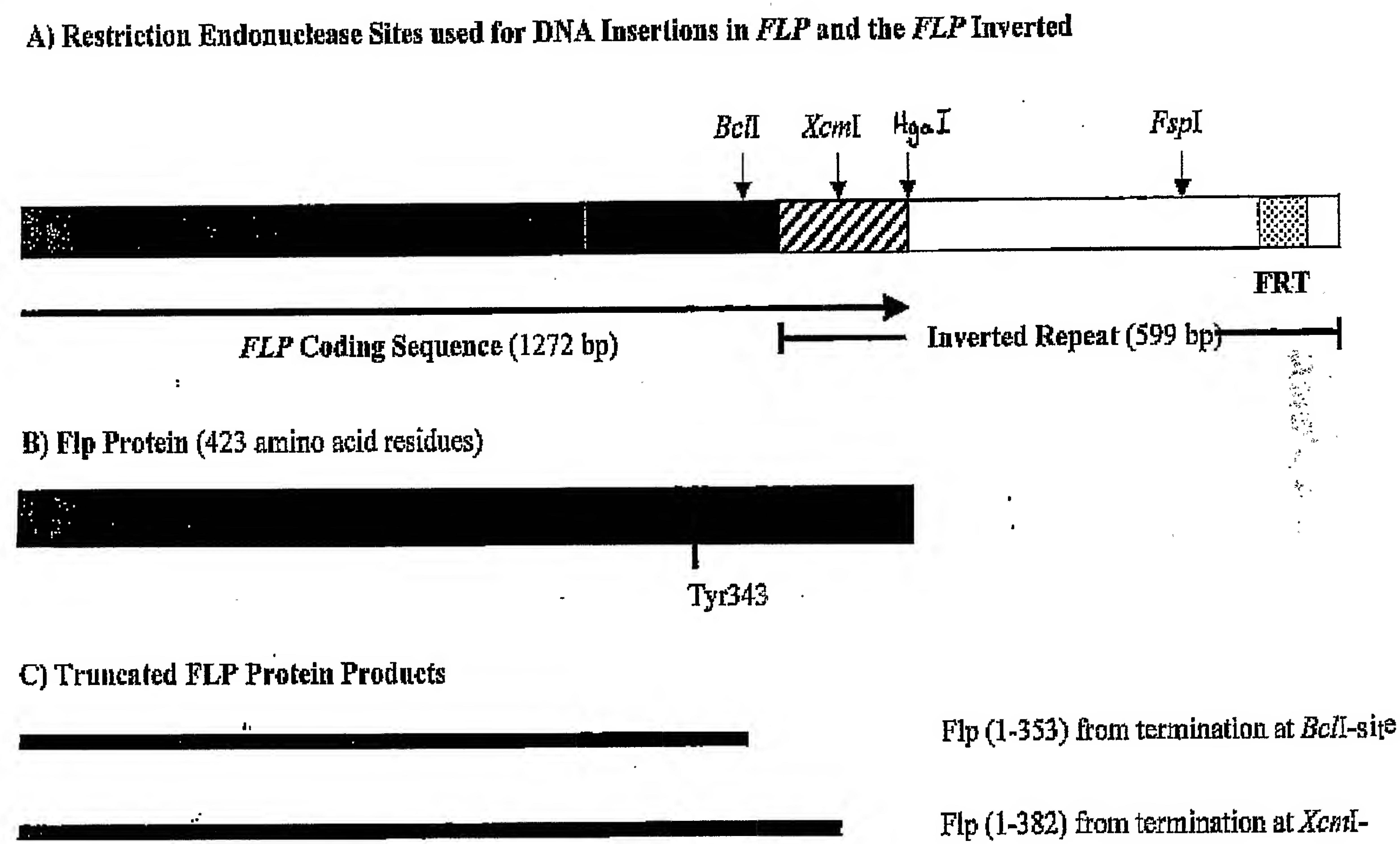
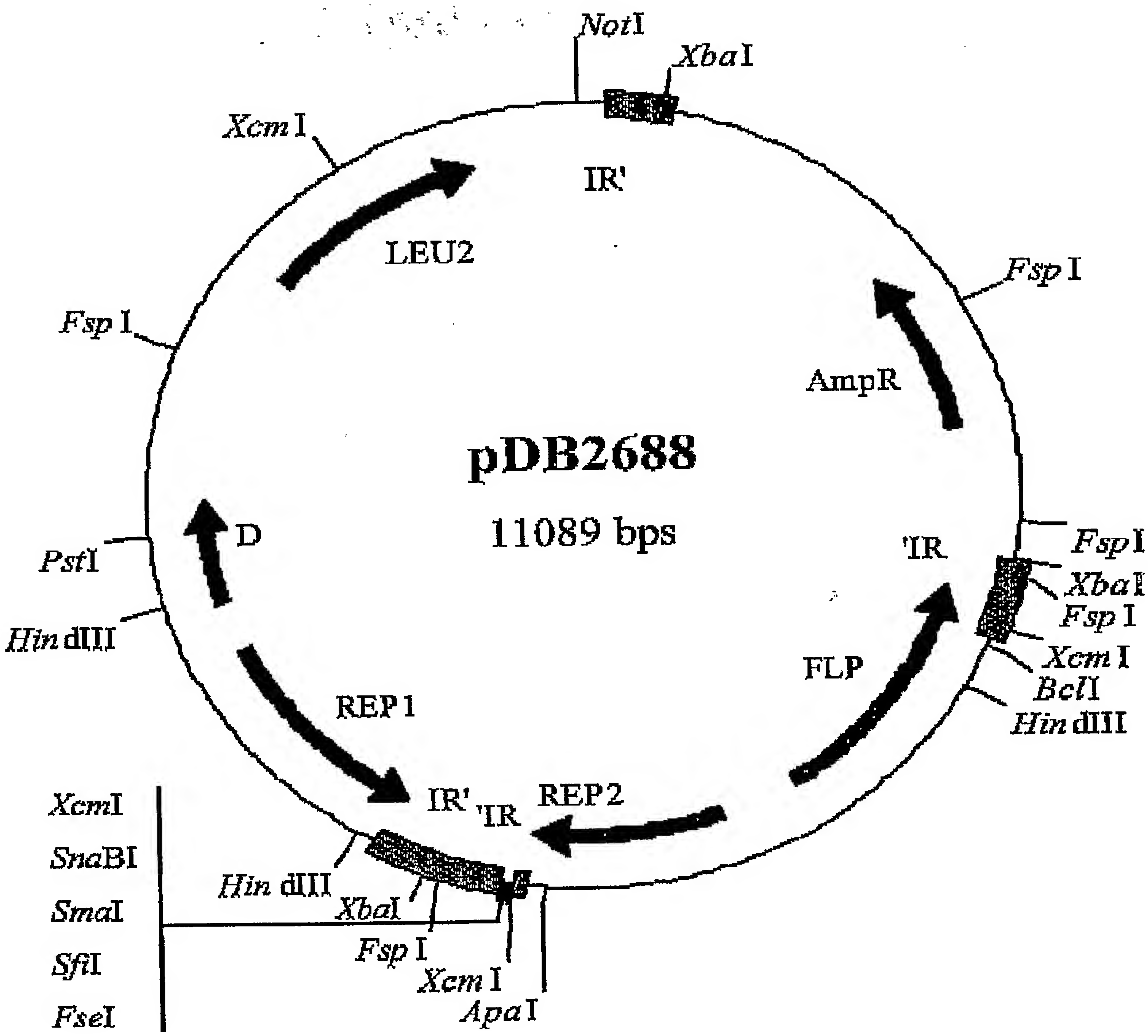

A) Restriction Endonuclease Sites used for DNA Insertions in *FLP* and the *FLP* Inverted
*Bcl*I
*Xcm*I
*Hga*I
*Fsp*I
FRT
FLP Coding Sequence (1272 bp)
Inverted Repeat (599 bp)
B) Flp Protein (423 amino acid residues)
Tyr343
C) Truncated FLP Protein Products
Flp (1-353) from termination at *Bcl*I-site
Flp (1-382) from termination at *Xcm*I-

Figure 4



NotI
XbaI
IR'
XcmI
LEU2
FspI
AmpR
FspI
pDB2688
11089 bps
PstI
D
HindIII
'IR
FspI
XbaI
FspI
XcmI
BclI
HindIII
FLP
REP1
IR'
'IR
REP2
XcmI
SnaBI
SmaI
SfiI
FseI
HindIII
XbaI
FspI
XcmI
ApaI

Figure 5

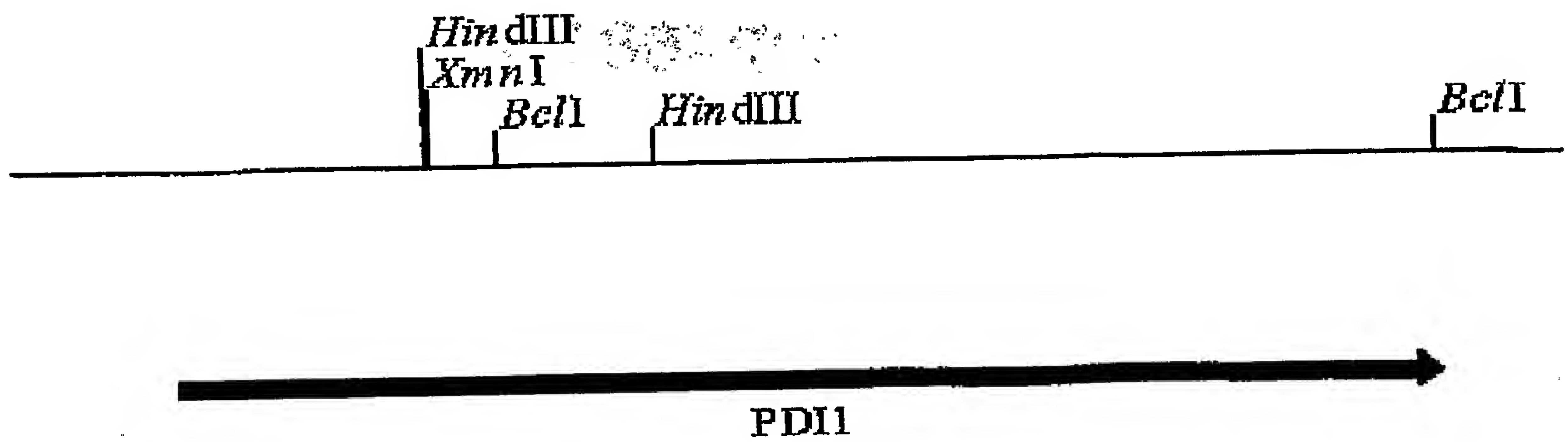
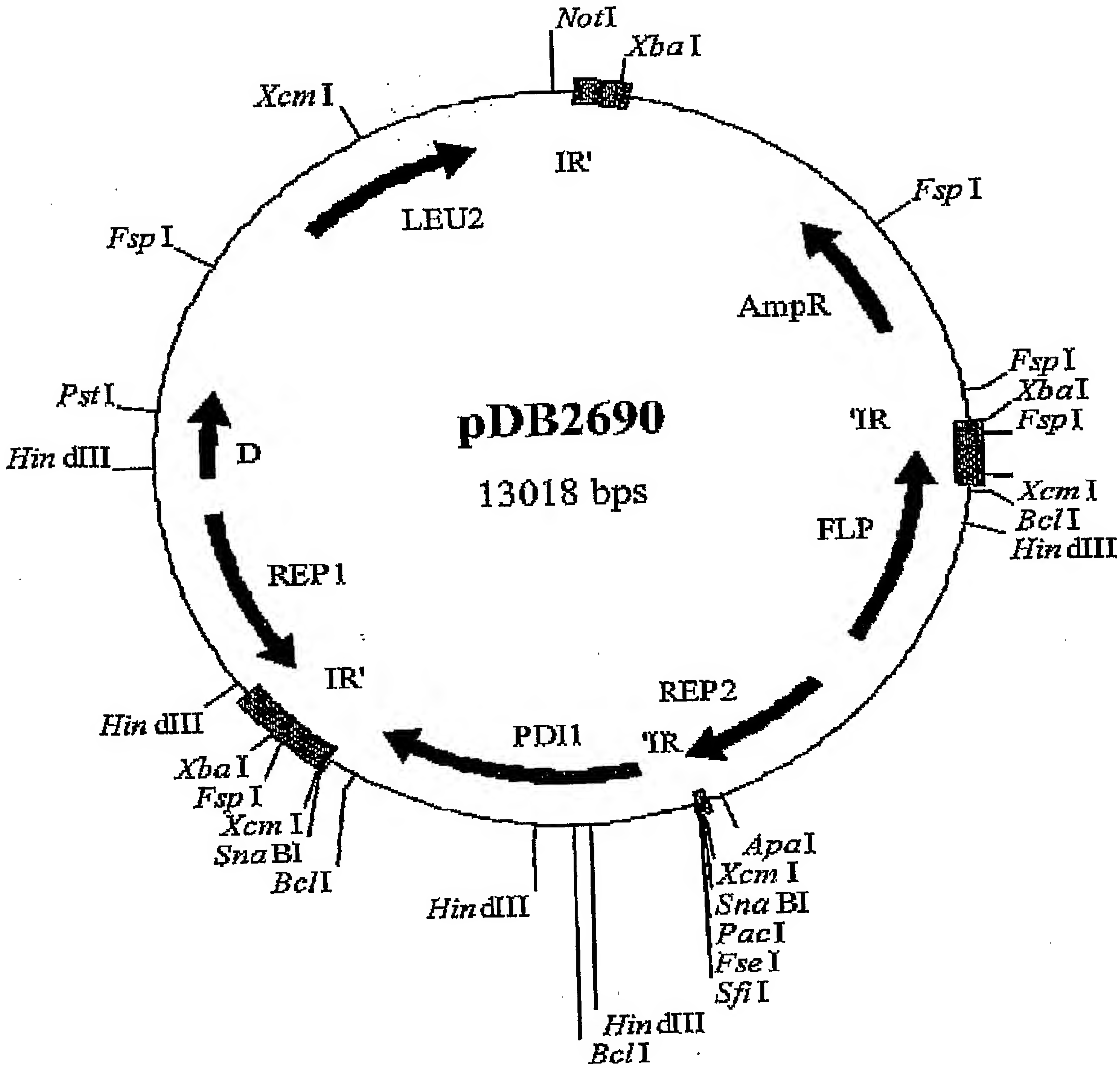


Figure 6



pDB2690
13018 bps
NotI
XbaI
XcmI
IR'
LEU2
FspI
AmpR
FspI
PstI
HindIII
D
FspI
XbaI
FspI
IR
XcmI
BclI
HindIII
FLP
REP1
IR'
REP2
HindIII
PDI1
IR
XbaI
FspI
XcmI
SnaBI
BclI
HindIII
ApaI
XcmI
SnaBI
PacI
FseI
SfiI
HindIII
BclI

Figure 7

mFL = modified HSA(pre)/MFα1(pro) fusion leader sequence

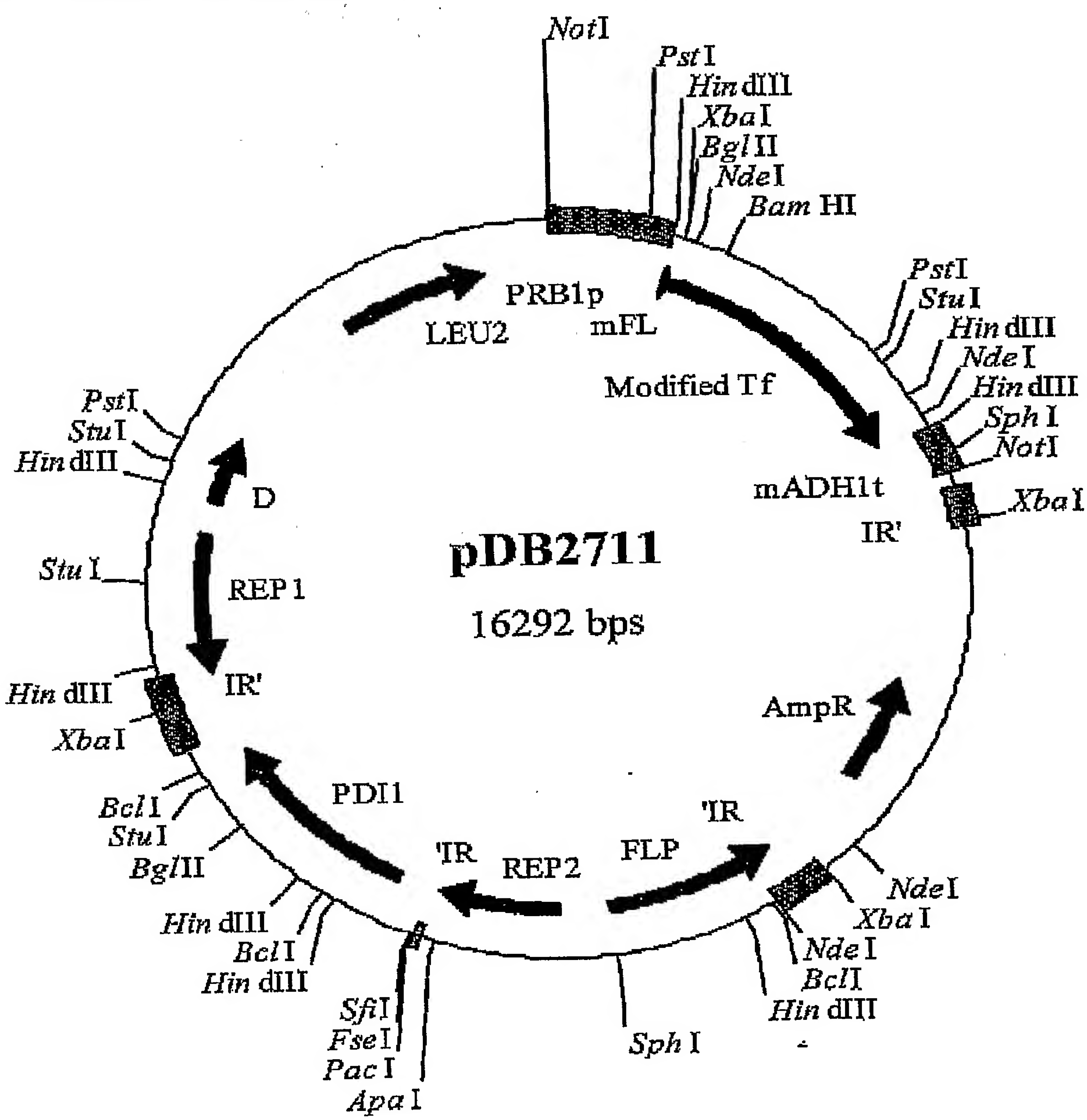


Figure 8

mFL = modified HSA(pre)/MFα1(pro) fusion leader sequence

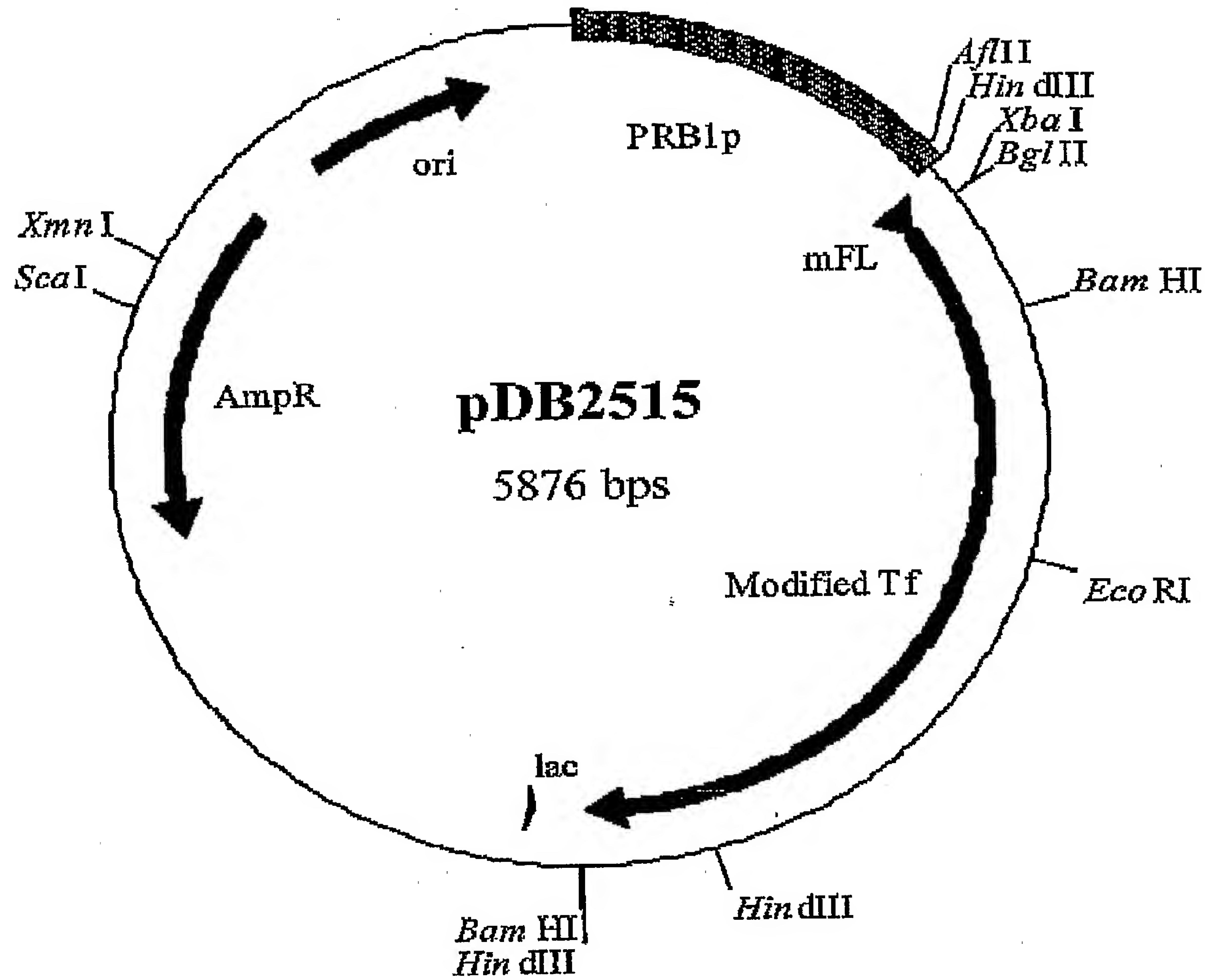


Figure 9

mHSA-pre = modified HSA-pre leader sequence

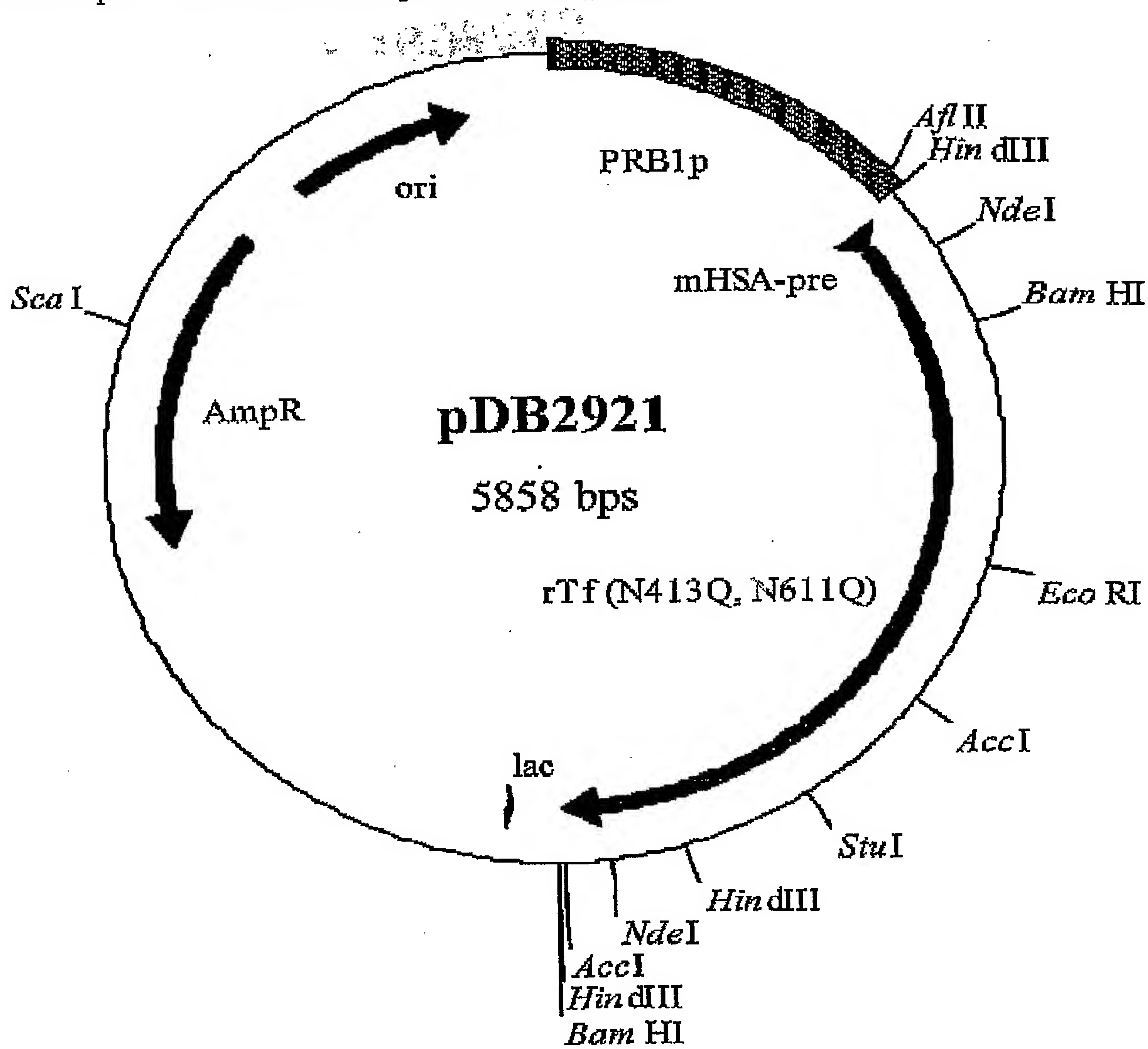


Figure 10

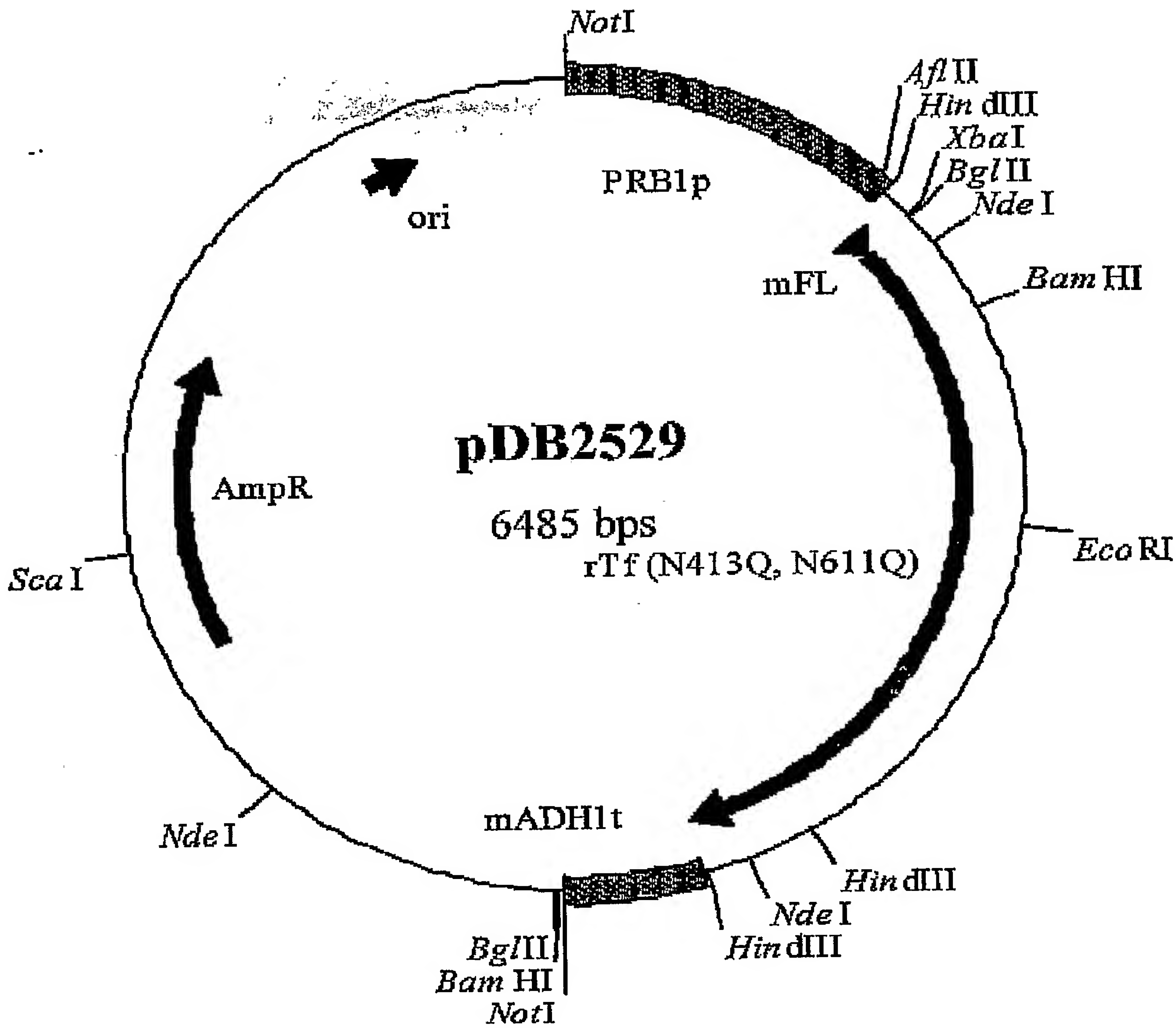

NotI
AflII
Hin dIII
XbaI
BglII
NdeI
BamHI
EcoRI
HindIII
NdeI
HindIII
BglII
BamHI
NotI
NdeI
ScaI
PRB1p
ori
mFL
AmpR
pDB2529
6485 bps
rTf (N413Q, N611Q)
mADH1t

Figure 11

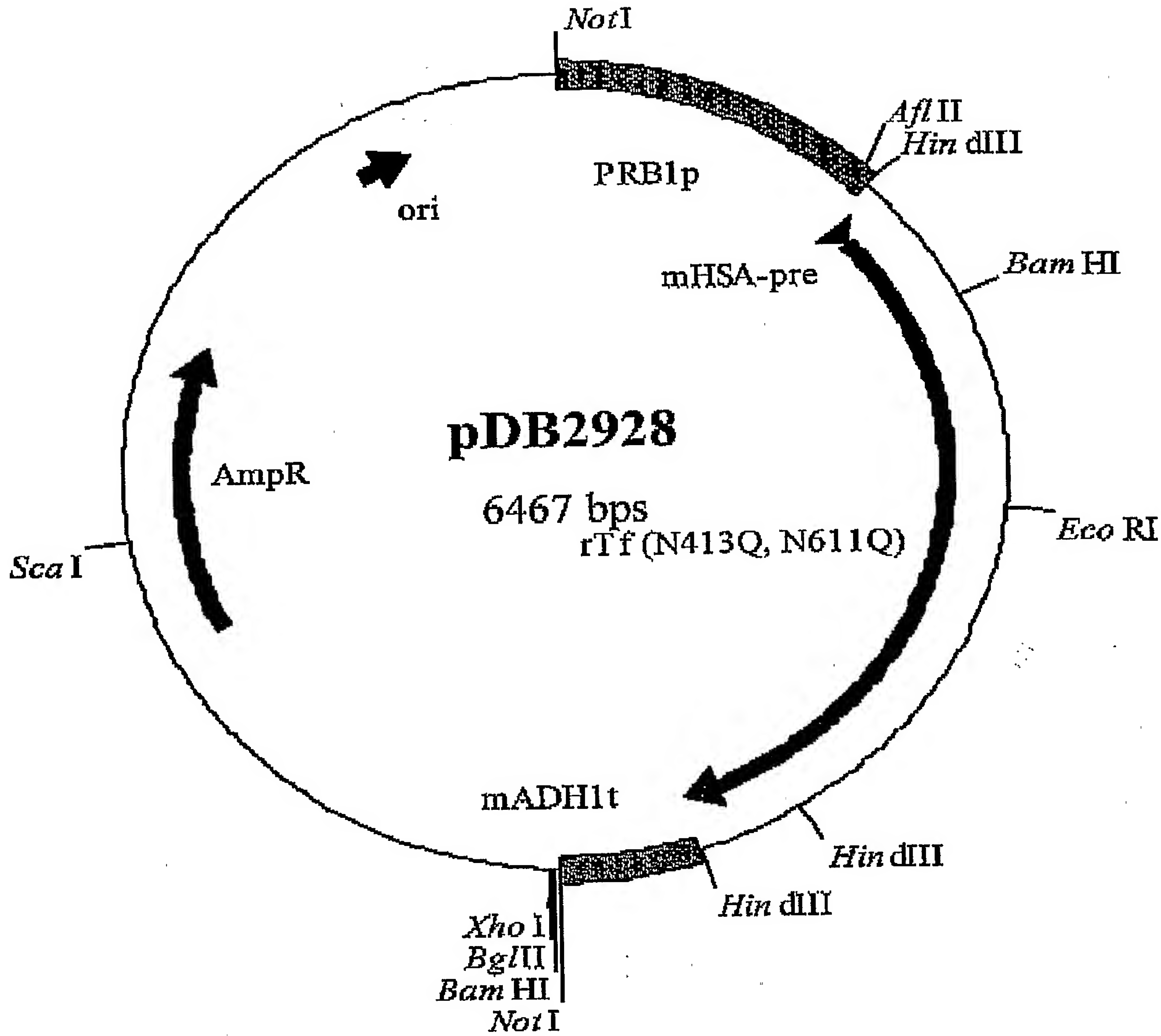
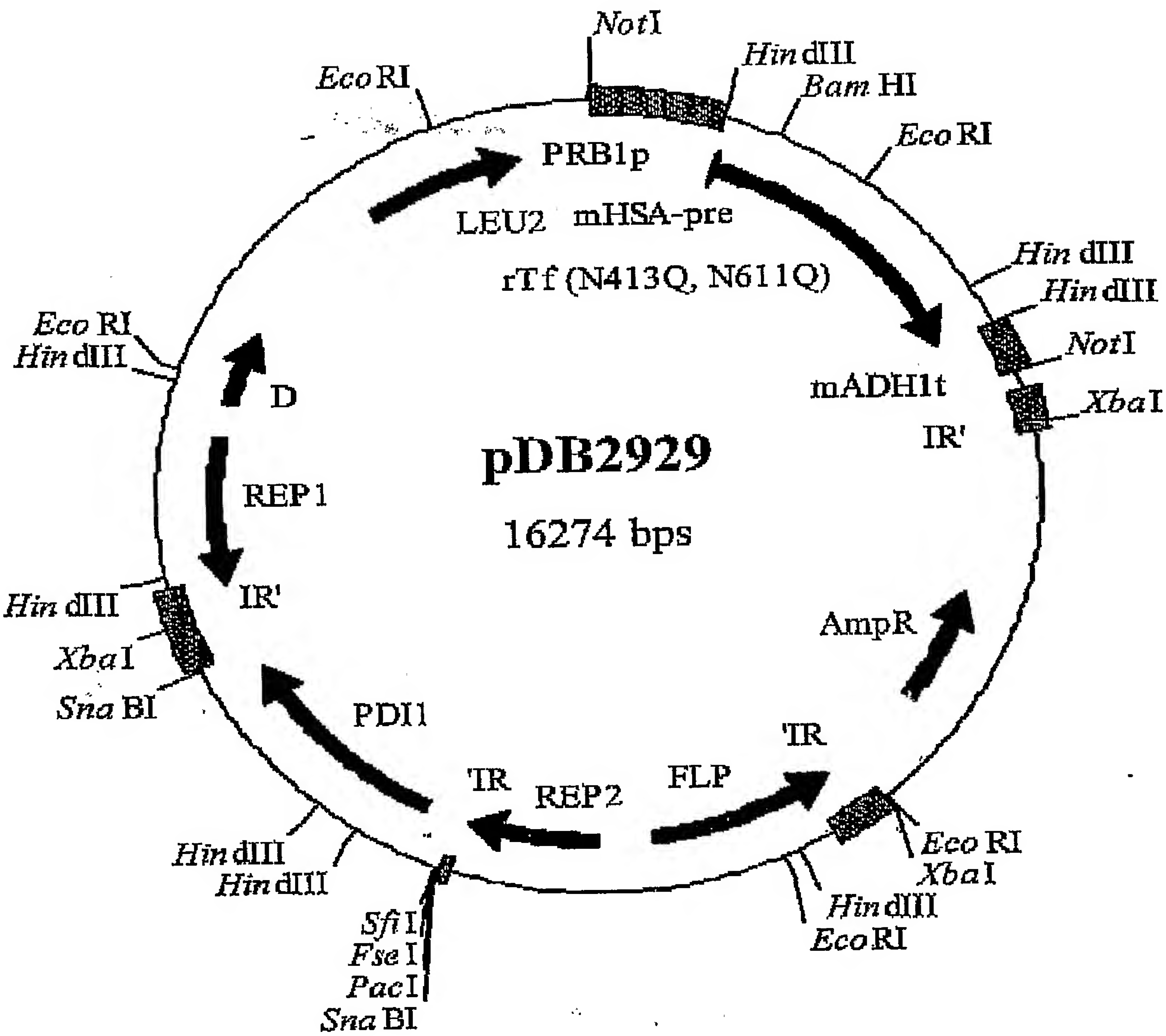
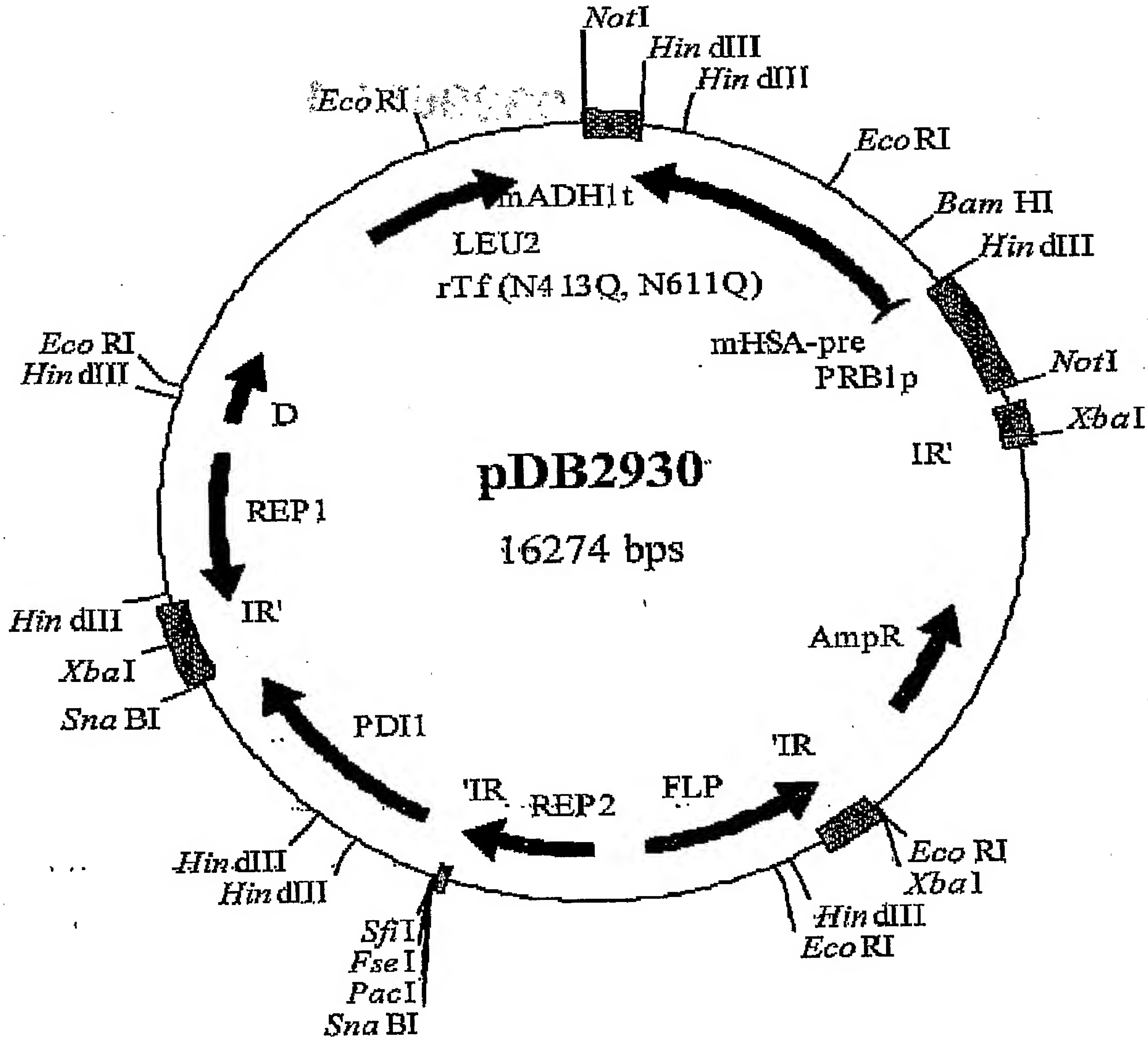

NotI
AflII
HindIII
PRB1p
ori
BamHI
mHSA-pre
pDB2928
6467 bps
rTf (N413Q, N611Q)
AmpR
EcoRI
ScaI
mADH1t
HindIII
HindIII
XhoI
BglII
BamHI
NotI

Figure 12



0089811 23-Dec-03 12:30

Figure 13



0089811 23-Dec-03 12:30

Figure 14

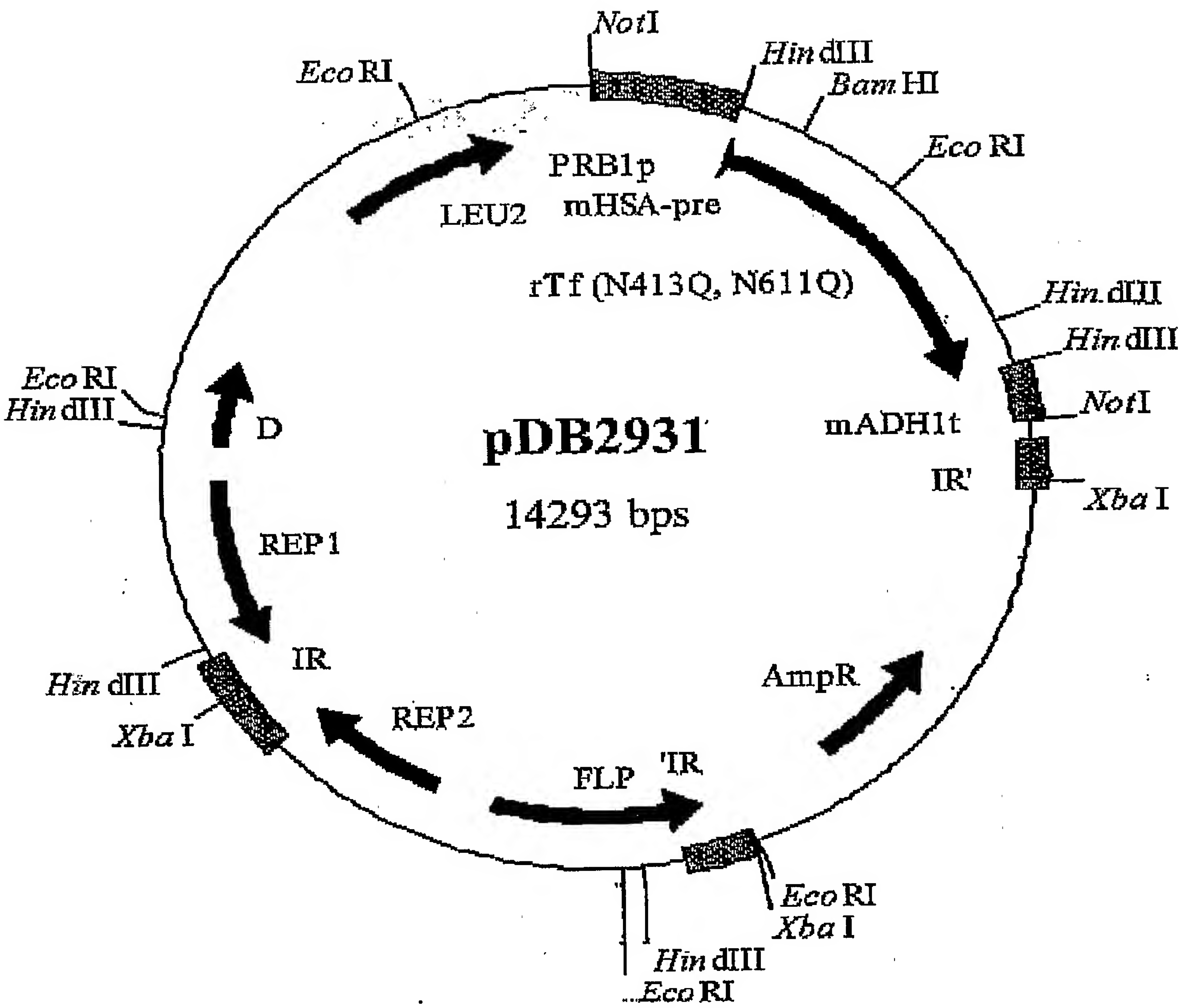
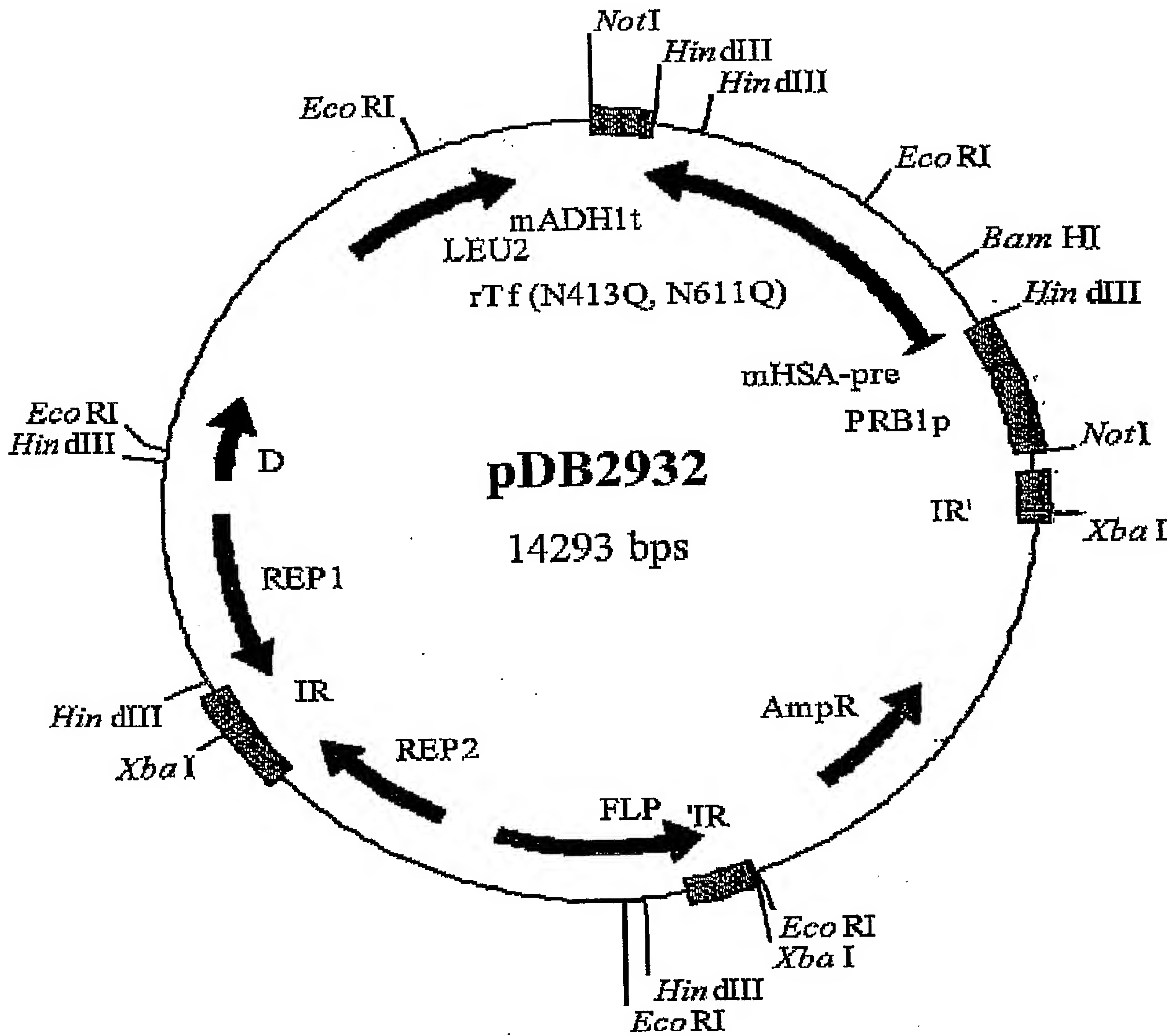
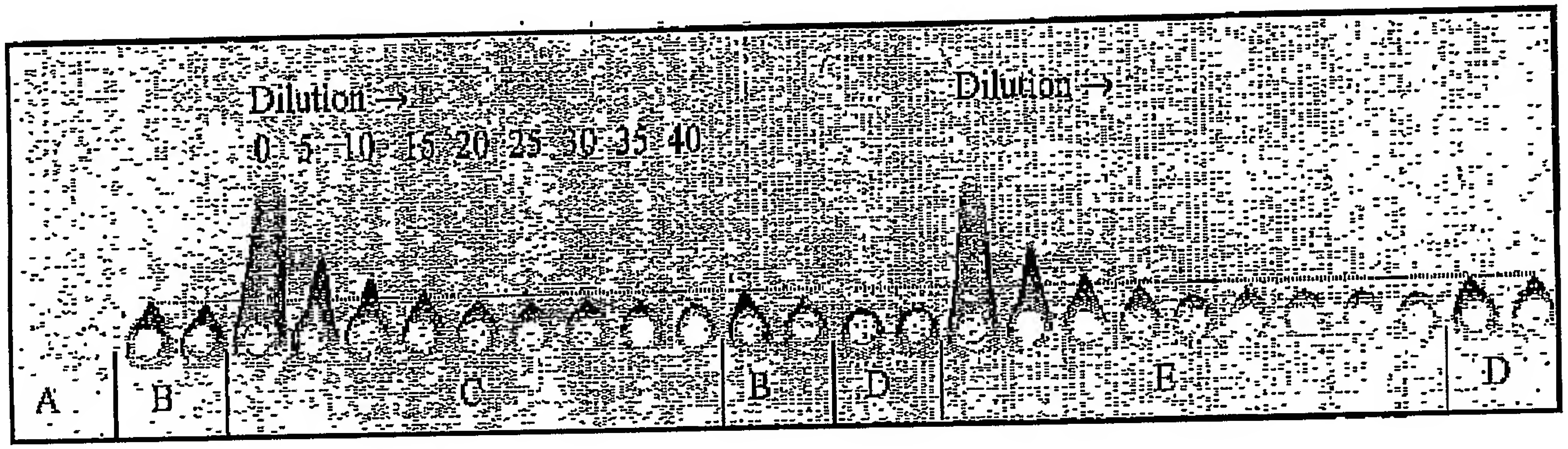

NotI
HindIII
BamHI
EcoRI
EcoRI
PRB1p
mHSA-pre
LEU2
rTf (N413Q, N611Q)
HindIII
HindIII
NotI
EcoRI
HindIII
D
pDB2931
14293 bps
mADH1t
IR'
XbaI
REP1
IR
HindIII
XbaI
REP2
AmpR
FLP
'IR
EcoRI
XbaI
HindIII
EcoRI

Figure 15



0089811 23-Dec-03 12:30

Figure 16



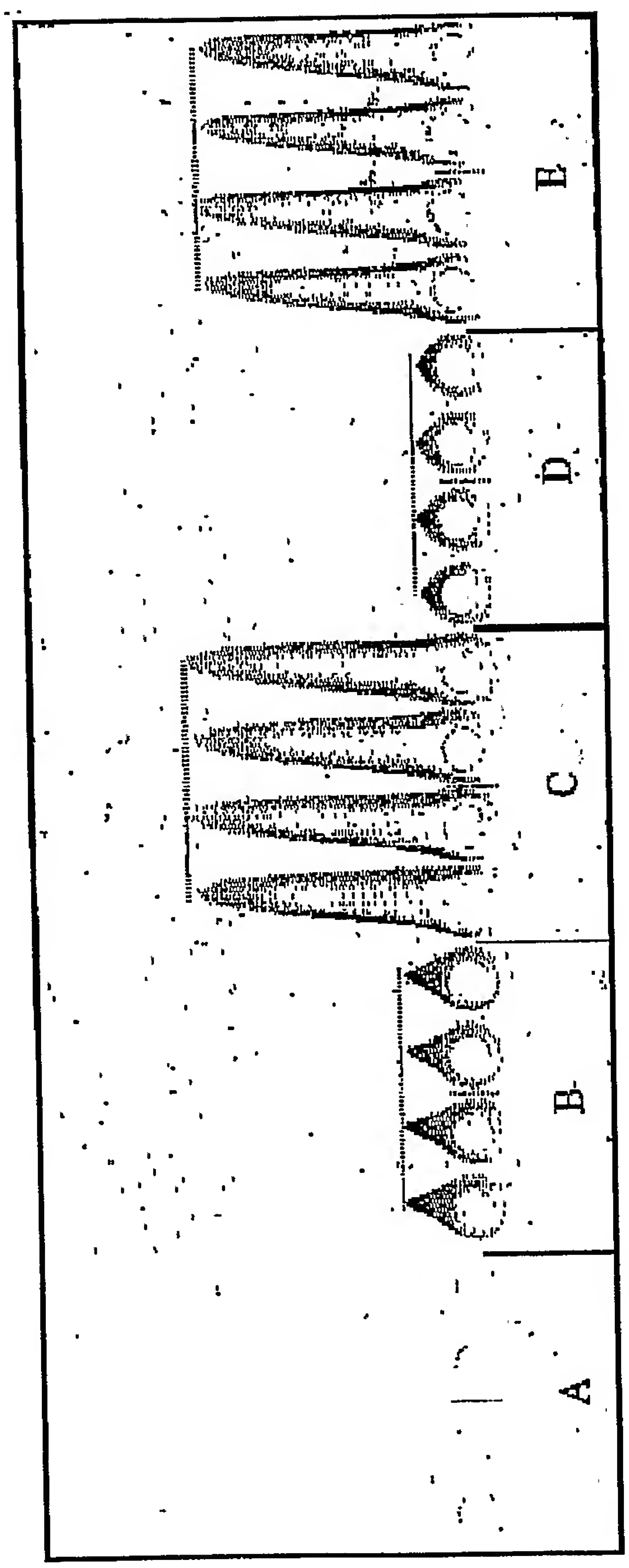


Figure 17

Figure 18

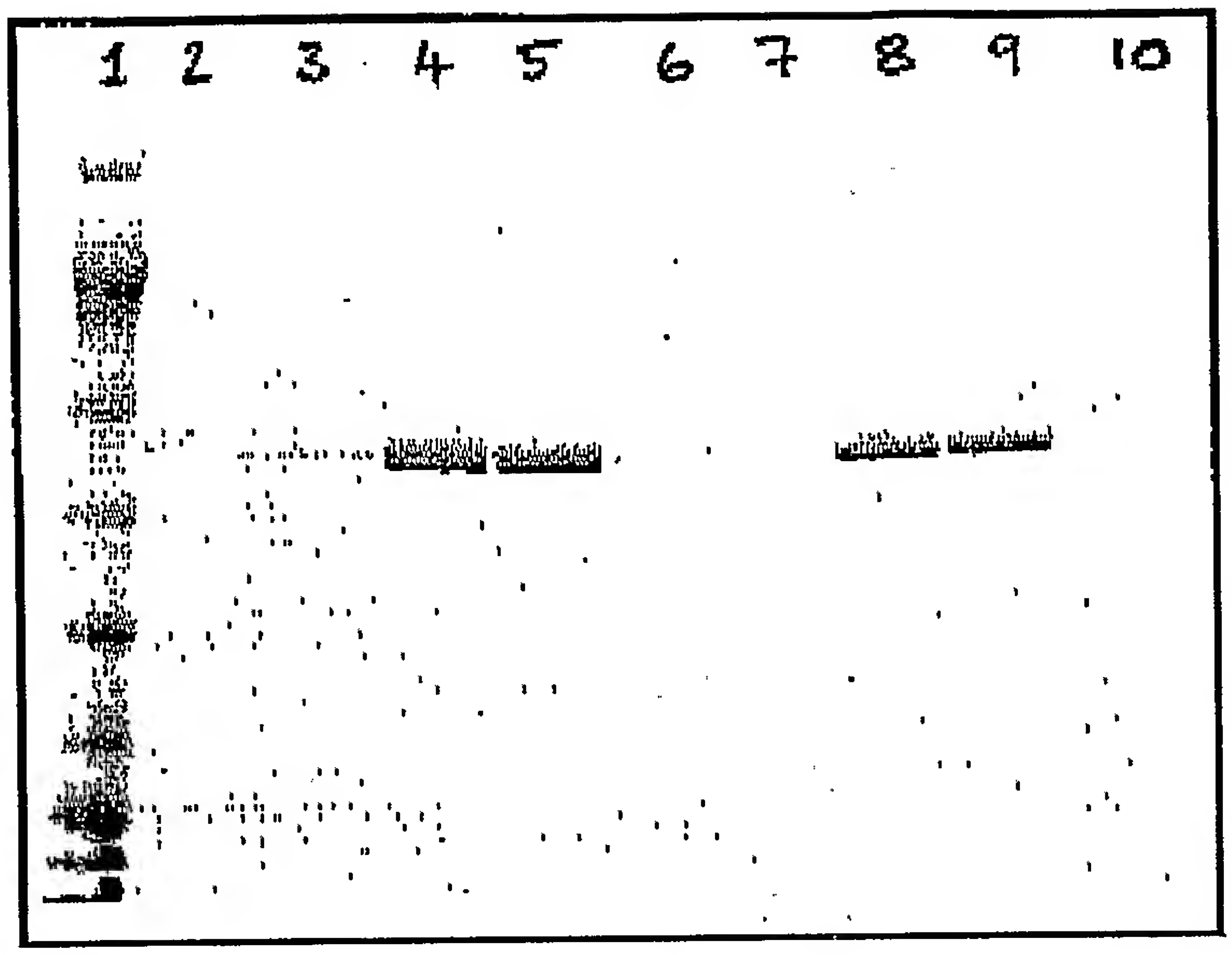


Figure 19

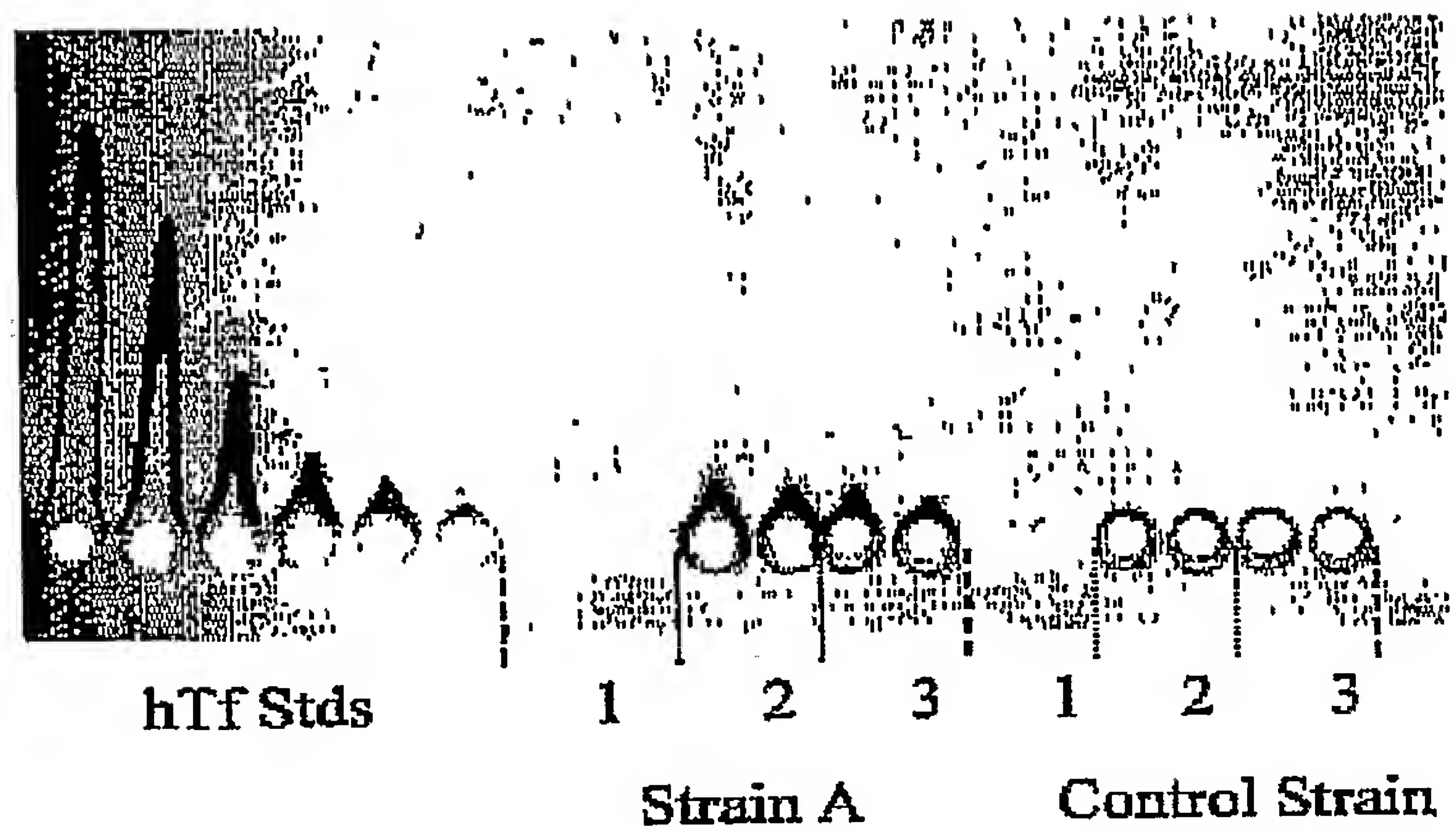


Figure 20

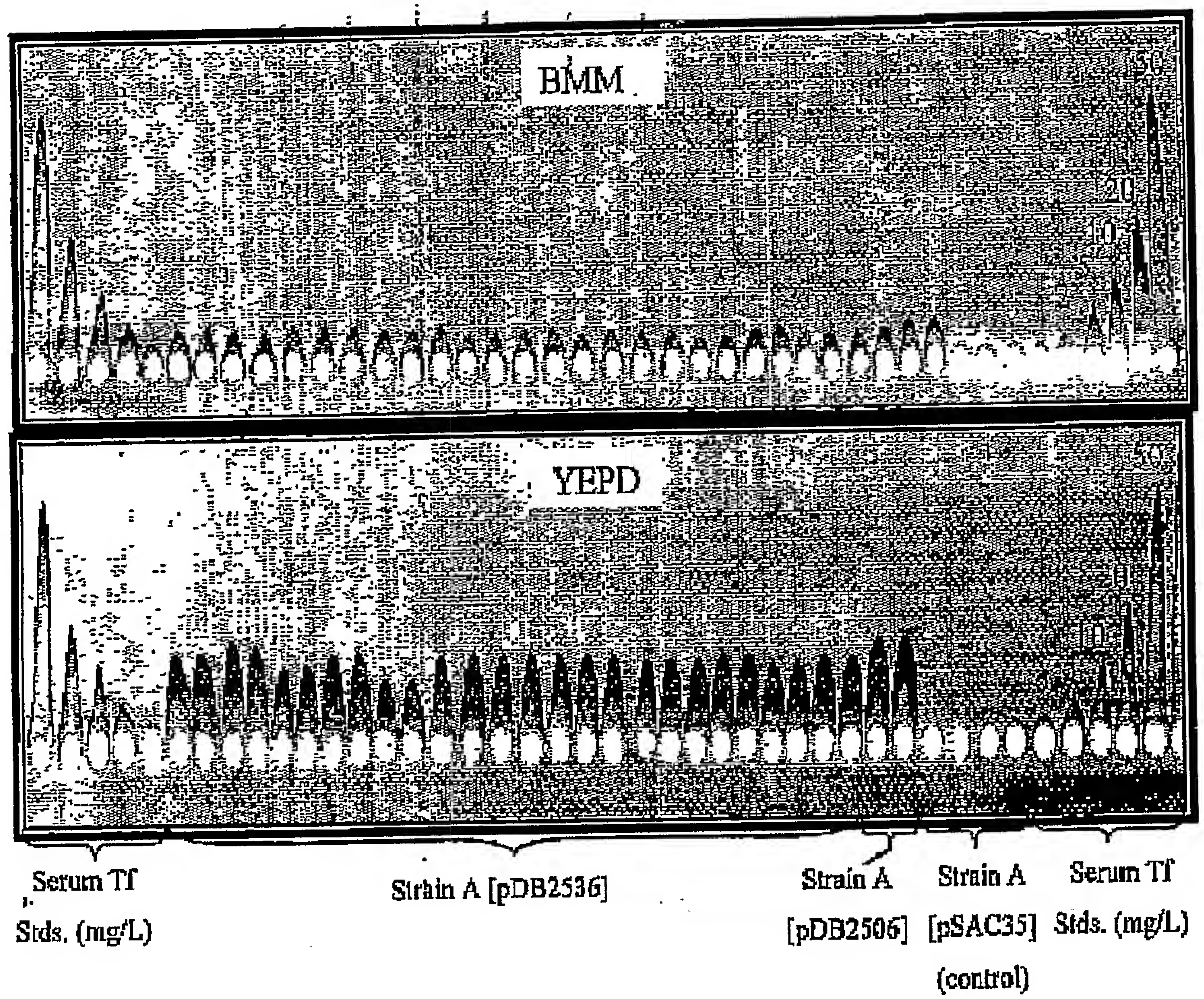
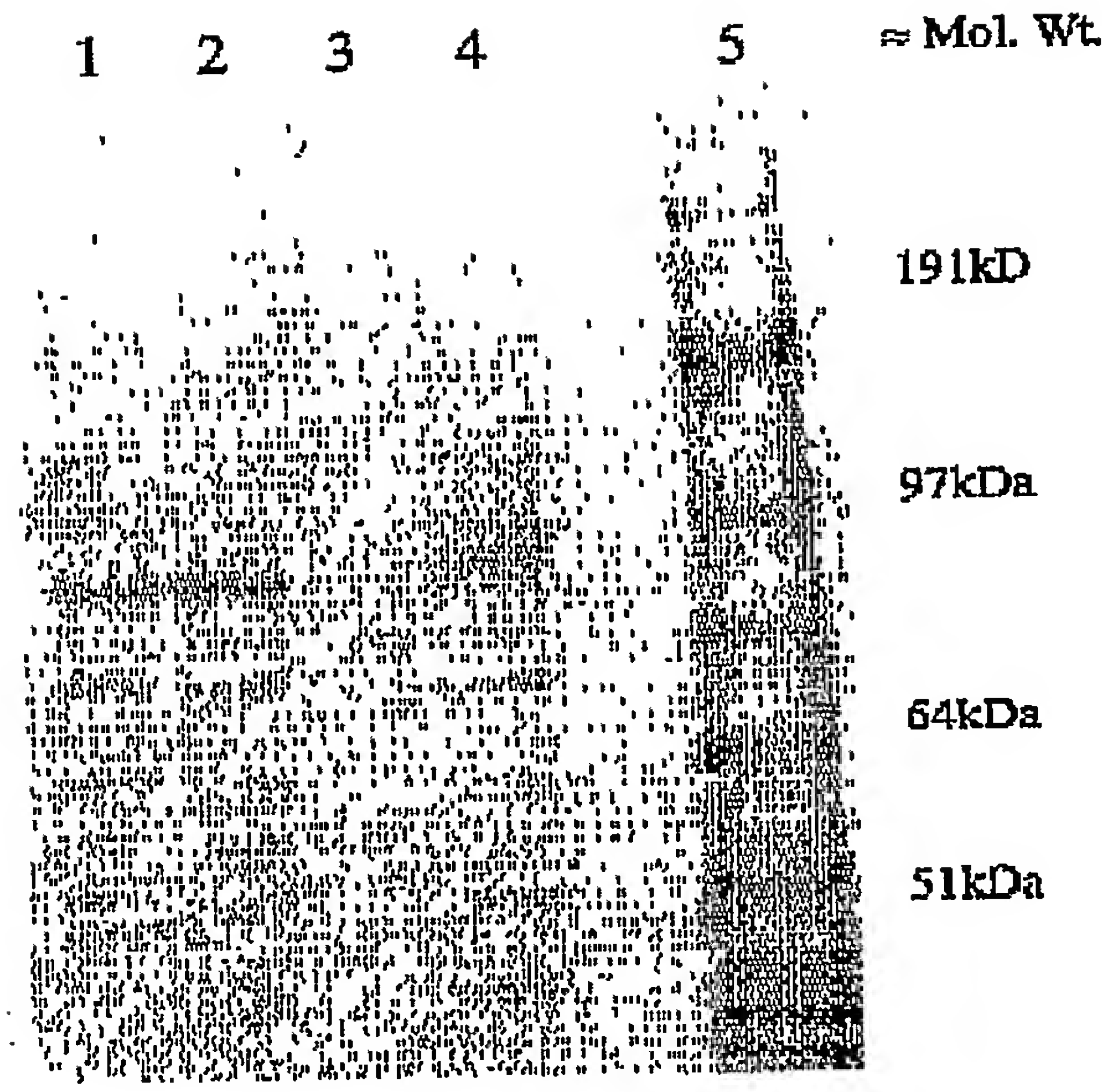
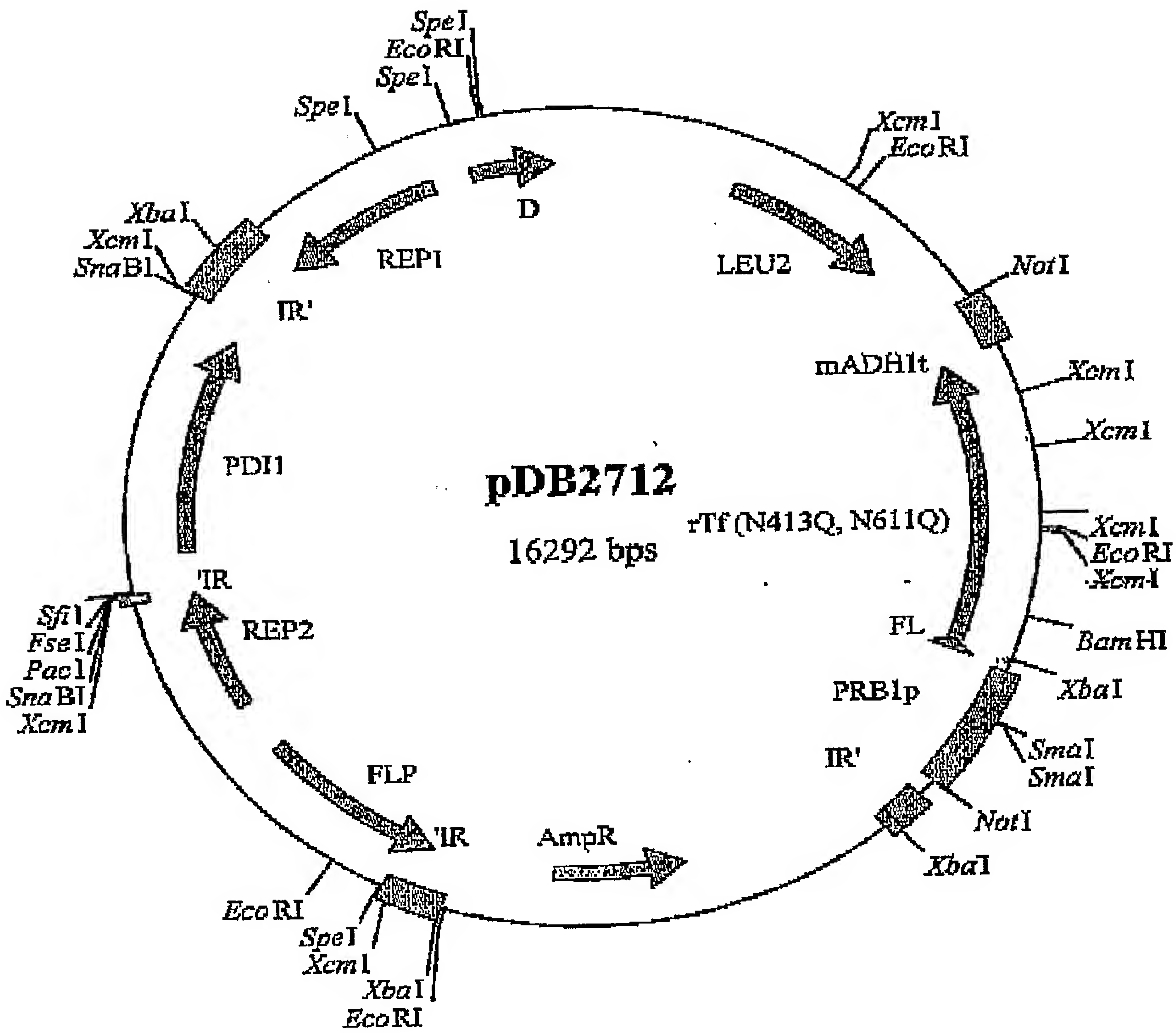

BMM
YEPD
Serum Tf Stds. (mg/L)
Strain A [pDB2536]
Strain A [pDB2506]
Strain A [pSAC35] (control)
Serum Tf Stds. (mg/L)

Figure 21



<u>**Figure 22**</u>



5

Figure 23

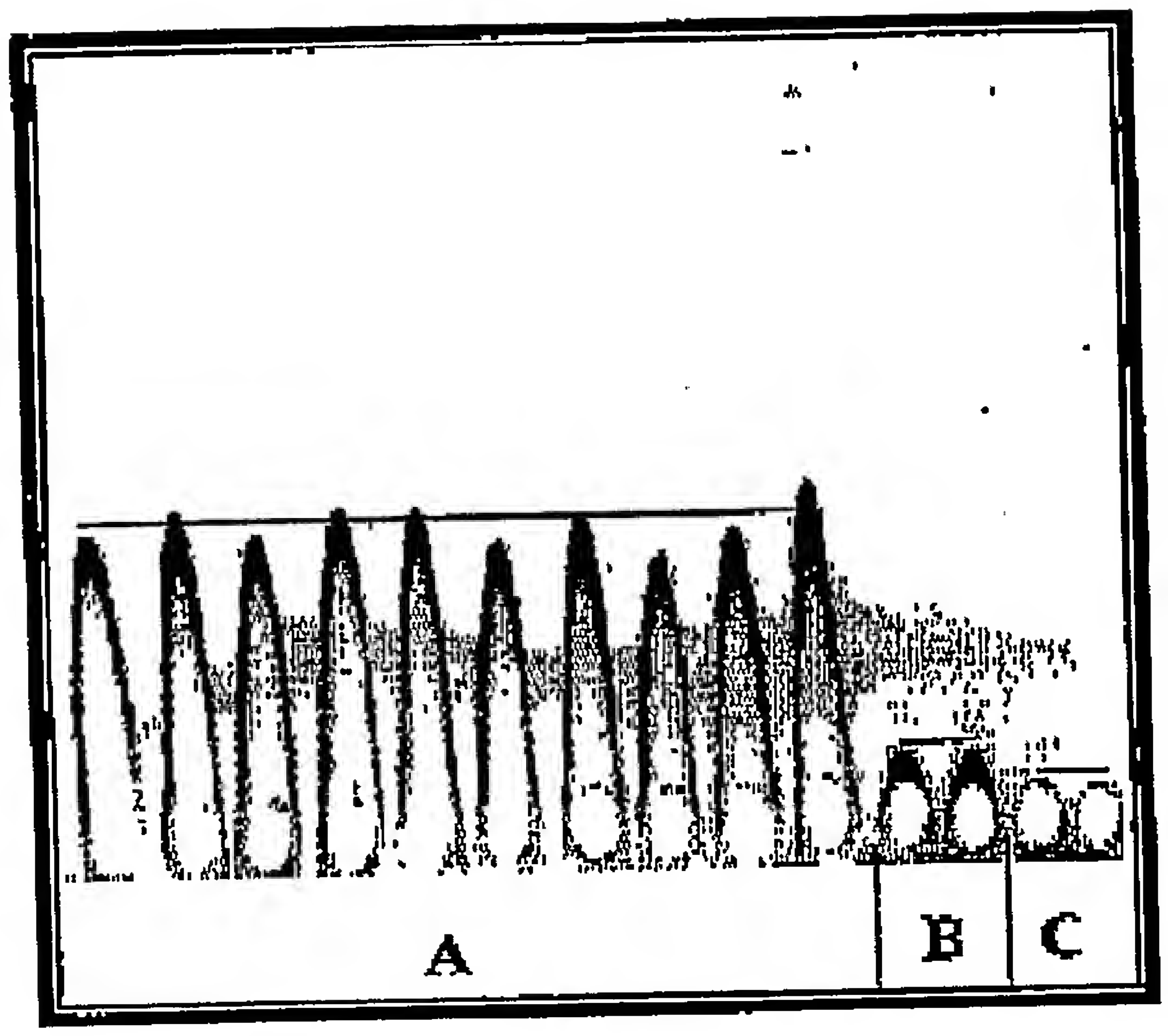


Figure 24

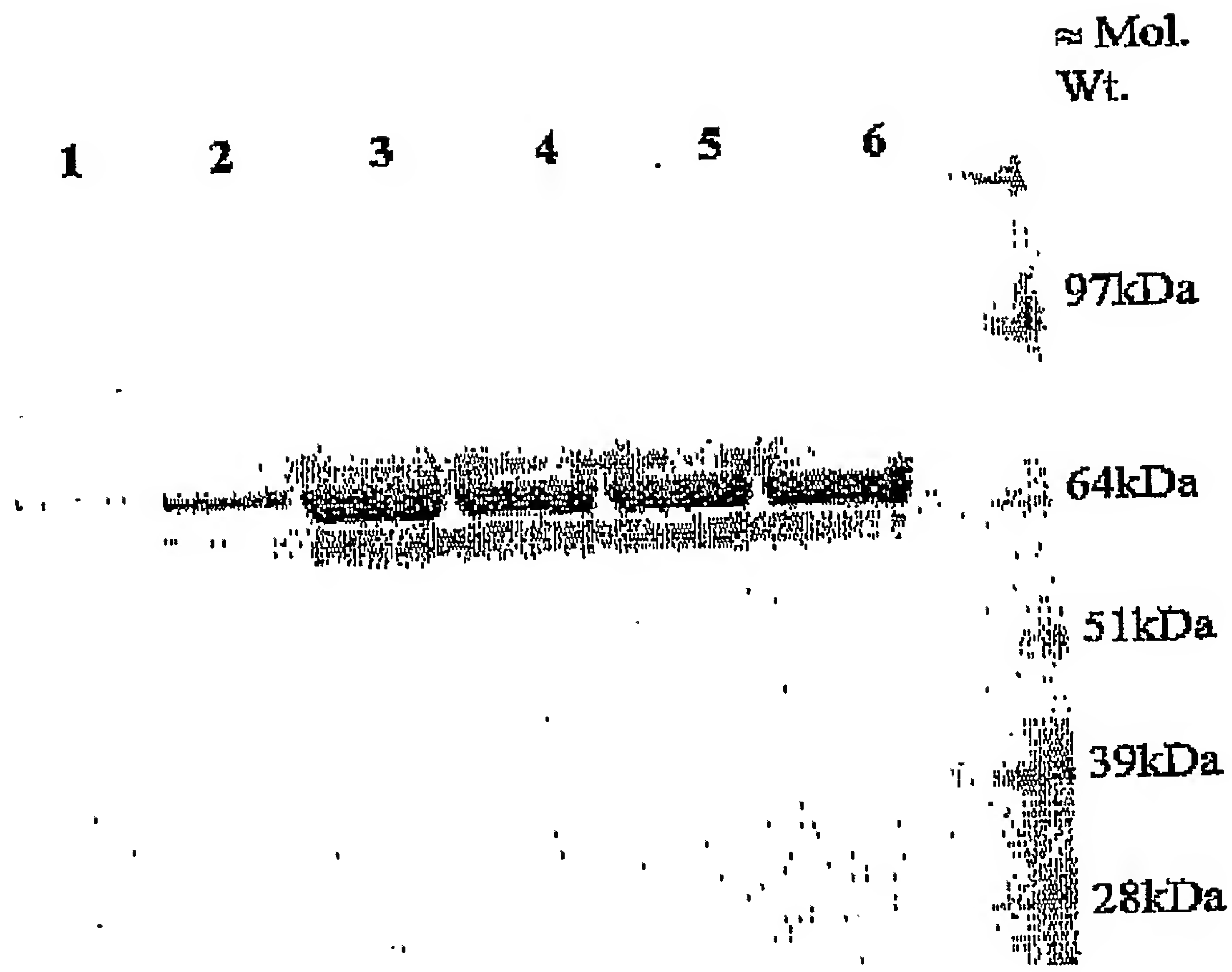
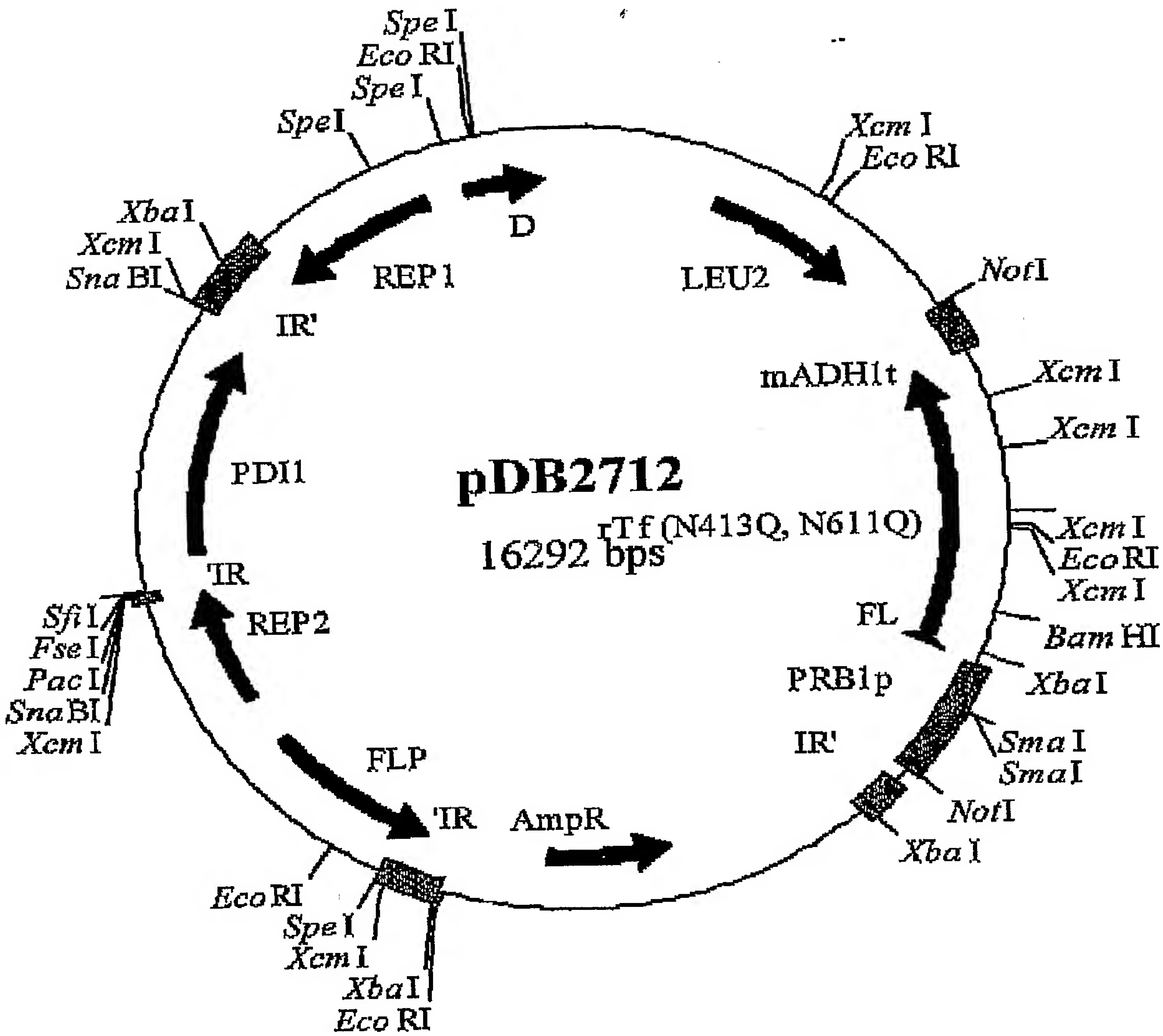


Figure 25


Spe I
Eco RI
Spe I
Spe I
Xcm I
Eco RI
Xba I
Xcm I
Sna BI
D
REP 1
LEU2
Not I
IR'
mADH1t
Xcm I
Xcm I
PDI1
pDB2712
16292 bps
rTf (N413Q, N611Q)
Xcm I
Eco RI
Xcm I
TR
Sfi I
Fse I
Pac I
Sna BI
Xcm I
REP2
FL
Bam HI
PRB1p
Xba I
IR'
Sma I
Sma I
FLP
Not I
TR
AmpR
Xba I
Eco RI
Spe I
Xcm I
Xba I
Eco RI